



Munich Personal RePEc Archive

# **Estimación del Modelo Probit Multivariante: Una Mejora**

Vargas, Martin

2003

Online at <https://mpra.ub.uni-muenchen.de/5241/>  
MPRA Paper No. 5241, posted 13 Oct 2007 UTC

Estimación del Modelo Probit Multivariante:  
Una Mejora.

Por

Martin H. Vargas Barrenechea

Tesis Para La obtención del Grado Academico de:  
LICENCIATURA EN ECONOMIA

LA PAZ BOLIVIA

2003

*A mi Madre, a quien le debo todo lo que soy y seré*

*En primer lugar agradezco al Lic. Fernando Rojas, Profesor tutor de esta tesis, por el inmenso apoyo y confianza que ha depositado en mi durante más de dos años, le agradezco también, por la invaluable asistencia en este documento como matemático y economista, y también por el valioso tiempo que dedico en revisar detalladamente a cada una de las más de 100 ecuaciones y a más de cinco complicadas demostraciones de este trabajo.*

*En segundo lugar deseo agradecer al Profesor Wilson Jiménez por sus críticas y recomendaciones, que me permitieron mejorar el presente documento.*

*Al Profesor Willian Greene por responder mis constantes preguntas y por sus valiosos comentarios, a Jurgen Doornik por la ayuda con la implementación de las derivadas numéricas en Ox, también a los profesores Shiddharta Chib, Arthur Lowell y Winfried Pohlmeier, también debo agradecer al Lic. Manuel Arana Director del Departamento de Economía, por el apoyo incondicional y generoso, que me brindo para poder defender la esta tesis.*

*Por ultimo y con igual importancia, quiero agradecer a los profesores Julio Humerez, Osvaldo Nina y Fernando Escobar por haberme dado la oportunidad de asistirlos en las clases de econometría y econometría dinámica en la UCB.*

# Índice general

<b>1. Introducción</b>	<b>4</b>
1.1. El modelo probit multivariante . . . . .	5
1.2. Antecedentes . . . . .	6
1.3. Objetivos . . . . .	7
<b>2. Análisis del modelo PMV</b>	<b>9</b>
2.1. Condición de Consistencia en Modelos PMV . . . . .	9
2.2. Condición de Identificación en el modelo PMV . . . . .	11
2.3. El logaritmo de verosimilitud del modelo PMV . . . . .	13
<b>3. Algunas herramientas de Computación y Optimización</b>	<b>15</b>
3.1. Métodos Numéricos de Optimización . . . . .	15
3.2. Estimación de probabilidades utilizando el algoritmo GHK . . . . .	18
3.3. Derivadas Numéricas . . . . .	20
<b>4. Estimación del modelo PMV</b>	<b>22</b>
4.1. Introducción . . . . .	22
4.2. Primeras Derivadas Analíticas . . . . .	22
4.3. Efectos Marginales . . . . .	32
<b>5. Aplicación</b>	<b>36</b>
5.1. La relación entre salud y pobreza: una aplicación . . . . .	37
5.2. Un Experimento de Monte Carlo . . . . .	43
<b>6. Conclusiones y Recomendaciones</b>	<b>47</b>
<b>A. Resultados de la estimación</b>	<b>52</b>

# Capítulo 1

## Introducción

Aunque las posibilidades *de construcción*, en modelos logísticos y modelos con estructura probit son inmensas, de la misma manera que, en la econometría clásica existe la posibilidad que variables continuas se encuentren correlacionadas, ya sea explícitamente, como es el caso de las ecuaciones simultáneas o implícitamente como es el caso de los sistemas SUR<sup>1</sup>, en el caso de modelos a variables binarias, existen combinaciones análogas y extensiones a casos multiecuacionales con estructuras variadas (multiprobit, multinomial logit y probit, tobit multivariante, etc.).

El modelo Probit Multivariante (PMV) es la extensión natural del modelo probit univariante, en el que los errores poseen una distribución normal multivariante estándar, algunas variables índice intervienen como endógenas y presentan varias formas de simultaneidad y causalidad. Los modelos PMV son de gran utilidad y su uso es cada vez más común y frecuente, dado que resuelven sesgos cuando se estudian variables discretas simultáneamente determinadas, con modelos univariantes y, además, permiten analizar algunos casos en los que algunas variables económicas sólo pueden ser definidas en forma discreta (crisis financiera, crisis cambiaria, etc). En los últimos años el tratamiento de datos en sección cruzada ha incrementado el uso de modelos PMV, bajo diferentes formas en sus supuestos básicos y estructurales, a fin de modelizar de forma más natural las relaciones económico sociales de los agentes económicos.

Los modelos PMV son no lineales, consecuencia de su estructura probit y, al igual que en el caso probit univariante, son estimados por el método de

---

<sup>1</sup>Seemingly Unrelated Regressions Model

Máximo Verosimilitud (MV) extendido, conocido como Máximo Verosimilitud con Información Completa (MVIC), dado que considera la información contenida en todas las ecuaciones de forma simultánea.

Sin embargo, la estimación de modelos PMV por el método MVIC es lenta y tediosa y, hasta hace algunos años, técnicamente imposible debido a que, la función de verosimilitud del modelo PMV, precisa la evaluación de integrales múltiples. Actualmente, los modelos PMV son estimados a través de los métodos del gradiente (BFGS, DFP, etc.) que, como veremos, precisan de las primeras derivadas del logaritmo de verosimilitud ( $\log L$ ). Estas son obtenidas de manera numérica, implicando un aumento exponencial en el número de cálculos requeridos, para obtener los estimadores Máximo verosímiles del modelo PMV. Mas aun, algunas variantes de los métodos del gradiente requieren de las segundas derivadas, incrementando el número de cálculos necesarios.

## 1.1. El modelo probit multivariante

El modelo probit univariante relaciona una variable no observada  $y^*$ , con una contraparte observada  $y$ , por la siguiente relación:

$$y_i = \begin{cases} 1 & \text{si } y_i^* > 0 \\ 0 & \text{si } y_i^* \leq 0 \end{cases} \quad (1.1)$$

donde se supone que:

$$y_i^* = x_i\beta + \varepsilon_i \quad (1.2)$$

$$\Pr(y_i = 1|x_i, \beta) = \Phi(y_i^*) = \Phi(x_i\beta + \varepsilon_i), \text{ y } \varepsilon_i \sim N(0, 1) \quad (1.3)$$

siendo  $\Phi$  la función de densidad acumulada (CDF) de la normal estándar,  $x_i$  es el vector de covariadas de dimensión  $1 \times k$  y  $\beta \in \mathfrak{R}^k$  es el vector de parámetros desconocidos. Estos son los supuestos estructurales probit.

A lo largo de este trabajo, adoptaremos la misma escritura, el superíndice \* representará una variable no observable y la misma variable, sin asterisco, representará su contraparte observable.

La extensión básica del modelo probit es el modelo PMV, que resulta de

ampliar el modelo probit univariante a un sistema de ecuaciones , es decir:

$$\begin{aligned} \begin{bmatrix} y_{i1}^* \\ \vdots \\ y_{iJ}^* \end{bmatrix} &= \begin{bmatrix} z_{i1}\beta_1 \\ \vdots \\ z_{iJ}\beta_J \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iJ} \end{bmatrix} \\ \varepsilon_i &\sim N_J(0, \Omega) \\ y_{ij} &= \begin{cases} 1 & \text{si } y_{ij}^* > 0 \\ 0 & \text{si } y_{ij}^* \leq 0 \end{cases} \end{aligned} \quad (1.4)$$

donde  $\Omega$  es simétrica y con diagonal igual a 1:  $\omega_{ii} = 1$  y como en cualquier modelo de ecuaciones simultáneas las covariadas  $z_{ij}$ , de dimensión  $1 \times k_j$ , pueden incluir elementos del vector de variables observadas  $\mathbf{y}_i = (y_{i1}, \dots, y_{iJ})'$ . Los vectores de parámetros desconocidos son  $\beta_1, \beta_2, \dots, \beta_J$ .y tenemos que  $\beta_j \in \mathfrak{R}^{k_j}$ .

Formalizando, sea  $Y_{ij}$  una variable aleatoria dicotómica, que asume los valores 1 ó 0 correspondientes a la observación  $i$  y a la ecuación  $j$ , tenemos entonces que  $\mathbf{y}_i$  es la colección de respuestas de las J variables del individuo  $i$ . Según el modelo PMV, la probabilidad de  $\mathbf{Y}_i = \mathbf{y}_i$ , condicionada sobre los parámetros, está dada por:

$$\Pr(\mathbf{Y}_i = \mathbf{y}_i) = \Pr(\mathbf{y}_i | \beta, \Omega) = \int_{A_J} \dots \int_{A_1} \phi_J(t | 0, \Omega) dt \quad (1.5)$$

donde  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iJ})'$ ,  $\phi_J(t|0; \Omega)$  es la función de densidad de la distribución normal estándar J-variada con vector de esperanzas 0 y matriz de correlación  $\Omega$ , el vector de parámetros ecuacionales es  $\beta' = (\beta'_1, \dots, \beta'_J)' \in \mathfrak{R}^k$ , con  $k = \sum k_j$ , donde  $k_j$  es el número de covariadas de la ecuación  $j$  y finalmente

$$A_j = \begin{cases} (-\infty, z_{ij}\beta_j) & \text{si } y_{ij} = 1 \\ (z_{ij}\beta_j, \infty) & \text{si } y_{ij} = 0 \end{cases}, \quad (1.6)$$

## 1.2. Antecedentes

La evolución de los modelos probit multivariante comenzó con el trabajo seminal de Zellner y Lee (1965), quienes propusieron un modelo bivariante de ecuaciones simultáneas con variables endógenas del tipo logit, estimado con un método de información limitada. Este modelo bivariante fue ampliado



y estimado con métodos de información completa por Amemiya (1974) sin coherencia ya que, las condiciones de consistencia, recién fueron desarrolladas después por Maddala y Lee (1976).

El método propuesto por Zellner y Lee fue generalizado para el caso multivariante por Amemiya(1977), para modelos de ecuaciones simultáneas no lineales. Esta generalización fue realizada utilizando las variables dependientes, tanto variables endógenas como *exógenas*, en su forma no observable y se denomina Ecuaciones Simultáneas Probit si las variables dependientes son binarias.

Aunque ya existía la base teórica suficiente para la estimación de modelos multivariantes de grado mayor a dos, solo a partir de 1996, varios programas integraron la estimación de modelos probit multivariantes de grado superior a dos, utilizando el algoritmo GHK<sup>2</sup> para el cálculo de las probabilidades. Aún así, son pocos los trabajos que utilizan la especificación probit multivariante con grado superior a dos, dada la complejidad de la función de verosimilitud y la lentitud con que estos modelos son ajustados.

Actualmente, existen pocos programas que incluyen las rutinas necesarias para la estimación del modelo probit multivariante. Por ejemplo, en STATA, se encuentran implementados los modelos probit, biprobit y triprobit. Un programa especializado en modelos de regresión con variables cualitativas es LIMDEP desarrollado por W. Greene, que incluye el cálculo de modelos probit multivariante hasta el grado 20 (probit 20-variante). Sin embargo, tanto STATA, LIMDEP y otros programas informáticos, utilizan métodos numéricos para el cálculo de las derivadas y el algoritmo GHK para la estimación de los modelos PMV de grado mayor a 2, por ende, las estimaciones de estos modelos son lentas.

### 1.3. Objetivos

Los principales objetivos de este trabajo son:

- Proporcionar una descripción general de los modelos PMV.
- Complementar la estimación MVIC de modelos PMV de orden mayor a dos, utilizando primeras derivadas analíticas del logaritmo de la función MV.

---

<sup>2</sup>Geweke (1989), Hajivassilou (1990), Keane (1994)

Los objetivos secundarios son:

- Obtener un instrumento teórico de estimación que incremente la velocidad y exactitud de los métodos actualmente utilizados.
- Verificar la consistencia del procedimiento comparando con los procedimientos establecidos en los programas STATA y LIMDEP

# Capítulo 2

## Análisis del modelo PMV

Los modelos probit de ecuaciones simultáneas son diferentes, tanto en relación con los modelos lineales de ecuaciones simultáneas como de modelos no lineales de ecuaciones simultáneas. En realidad, el modelo PMV puede ser considerado como un sistema de ecuaciones simultáneas con variables no observadas como dependientes y con variables observadas que pueden interactuar con otras variables exógenas, como covariadas. Sin embargo, estos modelos, además de la condición de identificación, deben satisfacer restricciones adicionales referentes a la existencia misma del modelo (consistencia) para cumplir con las propiedades de los modelos de probabilidad.

### 2.1. Condición de Consistencia en Modelos PMV

Escribiendo el modelo PMV para una sola observación de la siguiente forma<sup>1</sup>:

$$\mathbf{y}^* = \Gamma \mathbf{y} + B \mathbf{x} + \varepsilon \quad (2.1)$$

$\mathbf{y}^*$ ,  $\mathbf{y}$  y  $\varepsilon$  son vectores de dimensión  $J \times 1$ ;  $\mathbf{x}$  es  $K \times 1$ ,  $\Gamma$  es  $J \times J$ ; y  $B$  es  $J \times K$ , donde  $J$  es el número de ecuaciones,  $\mathbf{y}^* = (y_1^*, \dots, y_J^*)'$ ,  $\mathbf{y} = (y_1, \dots, y_J)'$  y la relación entre cada elemento de  $\mathbf{y}^*$  y sus contrapartes observadas es:

$$y_j = \begin{cases} 1 & \text{si } y_j^* > 0 \\ 0 & \text{si } y_j^* \leq 0 \end{cases} \quad j = 1, \dots, J \quad (2.2)$$

---

<sup>1</sup>Este capítulo omitirá el subíndice  $i$  es decir  $y_{i1}$  se escribirá como  $y_1$

**Lema 2 1** Sea  $y^*$  una variable no observable, con contraparte observable  $y$  definida como

$$y = \begin{cases} 1 & \text{si } y^* > 0 \\ 0 & \text{si } y^* \leq 0 \end{cases} \quad (2.3)$$

entonces la ecuación de la forma

$$y^* = \gamma y + \beta x - u \quad (2.4)$$

es lógicamente inconsistente a menos que  $\gamma = 0$ .

**Demostración (Maddala 1983).** Cuando  $y = 1$ , tenemos que  $\beta x + \gamma y > u \Rightarrow \beta x + \gamma > u$  y cuando  $y = 0$ , tenemos  $\beta x + \gamma y \leq u \Rightarrow \beta x \leq u$ , entonces,  $\Pr(y = 0) + \Pr(y = 1) = 1 - \Phi(\beta x) + \Phi(\beta x + \gamma)$ , es igual a uno solo si  $\gamma = 0$   $\square$

**Lema 2.2.** Sea el modelo bivalente

$$\begin{aligned} y_1^* &= \gamma_1 y_2 + \beta_1' x_1 - u_1 \\ y_2^* &= \gamma_2 y_1 + \beta_2' x_2 - u_2 \end{aligned} \quad (2.5)$$

el modelo es lógicamente consistente si  $\gamma_1 = 0$  ó  $\gamma_2 = 0$ .

**Demostración. (Maddala 1983).** Asumiendo que  $u_1$  y  $u_2$  son independientes, y sin pérdida de generalidad, supongamos que  $x_1 = x_2 = 0$ . Tenemos que:

$$\Pr(y_1 = 0, y_2 = 0) = [1 - \Phi(0)] [1 - \Phi(0)] \quad (2.6)$$

$$\Pr(y_1 = 0, y_2 = 1) = [1 - \Phi(\gamma_1)] \Phi(0) \quad (2.7)$$

$$\Pr(y_1 = 1, y_2 = 0) = \Phi(0) [1 - \Phi(\gamma_2)] \quad (2.8)$$

$$\Pr(y_1 = 1, y_2 = 1) = \Phi(\gamma_1) \Phi(\gamma_2) \quad (2.9)$$

la suma de estas probabilidades es:

$$1 + \Phi^2(0) - \Phi(0)\Phi(\gamma_1) - \Phi(0)\Phi(\gamma_2) + \Phi(\gamma_1)\Phi(\gamma_2)$$

la cual es diferente de uno, a menos que  $\gamma_1 = 0$  ó  $\gamma_2 = 0$   $\square$

**Condición General de Consistencia Interna (Schmith 1981).** Sea el sistema de ecuaciones PMV:

$$\mathbf{y}^* = \Gamma \mathbf{y} + B \mathbf{x} + \varepsilon \quad (2.10)$$

Como consecuencia del Lema 2.1 la matriz  $\Gamma$  debe tener ceros a lo largo de la diagonal y por el Lema 2.2 la matriz  $\Gamma$  es triangular inferior, expresada en alguna base, con ceros en la diagonal principal. Conjuncionando ambas condiciones, la condición general para la consistencia interna es: **Todos los menores principales de  $\Gamma$  deben ser iguales a cero**  $\square$

Esta condición, en el caso de los modelos SUR Probit, se cumple siempre ya que el modelo no incluye como variables explicativas a las contrapartes observables, por lo tanto  $\Gamma = 0$  y el modelo se escribe:

$$\mathbf{y}^* = B\mathbf{x} + \varepsilon \quad (2.11)$$

## 2.2. Condición de Identificación en el modelo PMV

Los modelos PMV son una clase especial de modelos no lineales, con variables no observables. El problema de identificación en modelos PMV es complejo dado que no posee una forma reducida conocida, por esta situación no es recomendable el uso de técnicas clásicas de ecuaciones simultáneas, este hecho fue demostrado en Maddala (1983), utilizando el siguiente modelo biprobit:

$$\begin{aligned} y_1^* &= x_1\beta_1 + \varepsilon_1 \\ y_2^* &= \gamma_{21}y_1 + x_2\beta_2 + \varepsilon_2 \end{aligned} \quad (2.12)$$

con  $\text{Var}(\varepsilon_1) = \text{Var}(\varepsilon_2) = 1$ . Este sistema no está identificado si  $\varepsilon_1$  y  $\varepsilon_2$  no son independientes y si  $x_2$  y  $x_1$  son constante, esto aunque ninguna combinación lineal de las dos ecuaciones genere una ecuación de la forma de la segunda. Aplicando estas condiciones (Maddala, 1983) el sistema se transforma en :

$$\begin{aligned} y_1^* &= \beta_1 + \varepsilon_1 \\ y_2^* &= \gamma_{21}y_1 + \beta_2 + \varepsilon_2 \end{aligned} \quad (2.13)$$

que posee 4 parámetros  $\beta_1$ ,  $\beta_2$ ,  $\gamma_{21}$  y  $\text{Cov}(\varepsilon_1, \varepsilon_2) = \delta$  y están disponibles sólo 3 probabilidades independientes, digamos  $p_{11}$ ,  $p_{01}$ ,  $p_{10}$  donde  $p_{ij} = \text{Pr}(y_1 = i, y_2 = j)$  para  $i, j = 0, 1$ , y la probabilidad  $p_{00}$  resulta de  $p_{00} = 1 - p_{11} - p_{01} - p_{10}$ . Por ello, solo es posible identificar 3 parámetros y el sistema está identificado si se impone una restricción sobre alguno de los parámetros  $\beta_1, \beta_2, \gamma_{21}, \delta$ , siendo la más común la condición de exclusión  $\beta_2 = 0$ .

Wilde (2000) demostró que la razón por la que el sistema 2.12 no es identificable es la elección de  $x_1$  y  $x_2$  dado que, al suponerlos constantes, solo existen 3 probabilidades independientes. Por otra parte, si suponemos que  $x_1 = x_2 = v$  donde  $v$  asume dos valores, por ejemplo 1 y 2, entonces el sistema es identificable dado que existen 4 parámetros y seis probabilidades independientes:  $p_{111}, p_{101}, p_{011}, p_{112}, p_{102}$  y  $p_{012}$ , donde  $p_{kmq}$  denota la probabilidad condicional  $\Pr(y_1 = k, y_2 = m | v = q)$ . Por lo tanto, existe suficiente variación en la información para identificar a los cuatro parámetros.

Consideremos el caso multivariante especificado al inicio del capítulo:

$$\mathbf{y}^* = \Gamma \mathbf{y} + B \mathbf{x} + \varepsilon \quad (2.14)$$

donde  $\mathbf{y}^*$ ,  $\mathbf{y}$ , y  $\varepsilon$  son vectores de dimensión  $J \times 1$ ;  $\mathbf{x}$  es  $K \times 1$ ;  $\Gamma$  es  $J \times J$ ; y  $B$  es  $J \times K$ , donde  $J$  es el número de ecuaciones,  $\mathbf{y}^* = (y_1^*, \dots, y_J^*)'$ ,  $\mathbf{y} = (y_1, \dots, y_J)'$  y la relación entre cada elemento de  $\mathbf{y}^*$  y sus contrapartes observadas es:

$$y_j = \begin{cases} 1 & \text{si } y_j^* > 0 \\ 0 & \text{si } y_j^* \leq 0 \end{cases} \quad j = 1, \dots, J \quad (2.15)$$

desarrollando el sistema 2.14 y aplicando la condición general de consistencia obtenemos el sistema genérico:

$$\begin{aligned} y_1^* &= x_1 \beta_1 + \varepsilon_1 \\ y_2^* &= \gamma_{21} y_1 + x_2 \beta_2 + \varepsilon_2 \\ y_3^* &= \gamma_{31} y_1 + \gamma_{32} y_2 + x_3 \beta_3 + \varepsilon_3 \\ &\vdots \\ y_J^* &= \gamma_{J1} y_1 + \gamma_{J2} y_2 + \dots + \gamma_{J,(J-1)} y_{J-1} + x_J \beta_J + \varepsilon_J \end{aligned} \quad (2.16)$$

donde  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_J)'$  es un vector con distribución Normal J-variante con  $\text{Var}(\varepsilon_j) = 1, j = 1, \dots, J$ , y  $E(\varepsilon) = 0$ . Wilde (2000) asume que  $x_j = 1$  para todo  $j$ , tomando esta restricción la primera ecuación del sistema posee solo un parametro desconocido ( $\beta_1$ ), la segunda ecuación del sistema dos parametros desconocidos ( $\gamma_{21}, \beta_2$ ), si tomamos en cuenta todos los parametros de las ecuaciones y los coeficientes de correlación el número total de parametros desconocidos del sistema es :

$$\sum_{n=1}^J n + \sum_{n=1}^J (n-1) = \frac{J(J+1)}{2} + \frac{(J-1)J}{2} = J^2 \quad (2.17)$$

y en este caso particular la función de verosimilitud contiene  $2^J - 1$  probabilidades independientes, desde que  $(2^J - 1) < J^2$  se cumple para  $1 < J < 5$ , los

modelos con dos, tres y cuatro ecuaciones no son identificables sin la existencia de regresores exógenos variantes, pero la existencia de regresores exógenos variantes soluciona el problema de identificación: Una variable exógena con dos valores conlleva 6 probabilidades independientes en el caso de dos ecuaciones, 14 en el caso de tres ecuaciones y a 30 probabilidades independientes en el caso de cuatro ecuaciones.

**Condición Débil de Identificación (Wilde 2000)** La existencia de un regresor exógeno variante en cada ecuación es suficiente para evitar problemas de identificación en modelos PMV.

## 2.3. El logaritmo de verosimilitud del modelo PMV

El modelo MVP consiste en un sistema de ecuaciones simultáneas con variables dependientes no observables, de la forma

$$\begin{aligned} y_{i1}^* &= z_{i1}\beta_1 + u_{i1} \\ y_{i2}^* &= z_{i2}\beta_2 + u_{i2} \\ &\vdots \\ y_{iJ}^* &= z_{iJ}\beta_J + u_{iJ} \end{aligned} \tag{2.18}$$

donde la relación entre variables observables  $y_{ij}$  y sus contrapartes no observadas  $y_{ij}^*$  es :

$$y_{ij} = \begin{cases} 1 & \text{if } y_{ij}^* > 0 \\ 0 & \text{if } y_{ij}^* \leq 0 \end{cases} \tag{2.19}$$

y los residuos  $\mathbf{u}_i = (u_{i1}, \dots, u_{iJ})'$  poseen una distribución normal J-variante  $N_J(0, \Omega)$ , donde  $\Omega = \{\omega\}_{ij}$  es la matriz de covarianzas de los residuos y  $\omega_{ii} = 1$ .

Tomaremos la siguientes convenciones,  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iJ})'$ ,  $\mathbf{y}_i = (y_{i1}, \dots, y_{iJ})'$ ,  $\phi_J(\cdot | \mu, \Sigma)$  representa la función de densidad normal J-variante con media  $\mu$  y matriz de varianzas y covarianzas  $\Sigma = \{\sigma\}_{km}$ . La distribución normal estándar J-variante corresponde al caso  $\mu = 0$ , y  $\sigma_{kk} = 1$  representándose su función de densidad por  $\phi_J(\cdot | 0, \Omega)$ , la función  $\Phi_J(l_1, \dots, l_J, \Omega)$  representa la distribución acumulada de la normal estándar J-variante con límites  $l_j$  y matriz de correlación  $\Omega$  y en el caso bivariante se tiene la equivalencia: .y tomaremos las siguientes equivalencias notacionales:

1.  $\phi_1(\cdot | 0, 1) \equiv \phi_1(\cdot)$
2.  $\Phi_2(l_1, l_2, \Omega) \equiv \Phi_2(l_1, l_2, \omega_{12})$

La probabilidad  $\Pr(\mathbf{y}_i | \beta, \Omega)$  es:

$$\Pr(\mathbf{y}_i | \beta, \Omega) = \int_{A_J} \dots \int_{A_1} \phi_J(t | 0, \Omega) dt, t \in \mathfrak{R}^J \quad (2.20)$$

con

$$A_j = \begin{cases} (-\infty, z_{ij}\beta_j) & \text{si } y_{ij} = 1 \\ (z_{ij}\beta_j, \infty) & \text{si } y_{ij} = 0 \end{cases}, \quad (2.21)$$

donde  $\beta_j \in \mathfrak{R}^{k_j}$  es un vector de parámetros desconocidos y  $\beta' = (\beta'_1, \dots, \beta'_J)' \in \mathfrak{R}^k$ , con  $k = \sum k_j$ , es el vector de parámetros ecuacionales.

La ecuación 2.20 puede ser reescrita como:

$$\Pr(\mathbf{Y}_i = \mathbf{y}_i) = \Phi_J(w_{i1}, \dots, w_{iJ}, R_i) \quad (2.22)$$

donde  $w_{ij} = q_{ij}z_{ij}\beta_j$  con  $q_{ij} = 2y_{ij} - 1$  y  $R_i = \{\rho_i\}_{km}$  es la matriz de correlación con  $\rho_{ikm} = q_{ik}q_{im}\omega_{km}$ .

Por lo tanto, el **logaritmo de verosimilitud**  $\log L$  del modelo PMV correspondiente es:

$$\log L = \sum_{i=1}^n \log \Phi_J(w_{i1}, \dots, w_{iJ}, R_i) \quad (2.23)$$

y utilizando  $\mathbf{w}_i = (w_{i1}, \dots, w_{iJ})$ , se obtiene la expresión siguiente:

$$\log L = \sum_{i=1}^n \log \Phi_J(\mathbf{w}_i, R_i) \quad (2.24)$$

los estimadores MVIC son los estimadores  $\hat{\beta}$  y  $\hat{\Omega}$  que maximizan la ecuación 4.1, siendo que este modelo es no lineal es necesario el uso de algoritmos numéricos de búsqueda para encontrar estos estimadores y, además, es necesario evaluar probabilidades J variantes que, a partir del caso trivariante, son complejas y aproximadas por el algoritmo GHK. Estas herramientas son desarrolladas en el siguiente capítulo.



# Capítulo 3

## Algunas herramientas de Computación y Optimización

Dado que los modelos PMV son no lineales se requiere el uso de métodos numéricos de optimización para la maximización del  $\log L$  y métodos de simulación de momentos para el cálculo de probabilidades superiores al caso bivariante. Esta sección provee una descripción básica acerca de los métodos numéricos aplicados a la estimación por MV y sobre el cálculo de derivadas numéricas. También describe el método GHK para el cálculo de probabilidades en el caso de la distribución normal multivariante.

### 3.1. Métodos Numéricos de Optimización

Sea  $C$  una aplicación de  $\mathfrak{R}^p$  en  $\mathfrak{R}$ , de clase  $C^2(\Theta)$ , donde  $\Theta \subseteq \mathfrak{R}^p$  es el espacio de posibles parámetros, Sea  $\theta$  un punto de este espacio y sea  $\theta_0(p \times 1) \in \Theta$  el verdadero valor que se desea estimar. Si se define un estimador de  $\theta_0$  basado en una muestra de observaciones de tamaño  $n$  por :

$$\hat{\theta} = \arg \min_{\theta \in \Theta} C(\theta) \quad (3.1)$$

además, sea el gradiente:

$$\nabla C = \frac{\partial C}{\partial \theta} = \begin{bmatrix} \frac{\partial C}{\partial \theta_1} \\ \vdots \\ \frac{\partial C}{\partial \theta_p} \end{bmatrix} \quad (p \times 1) \quad (3.2)$$

y la matriz hesiana:

$$\nabla^2 C = \frac{\partial^2 C}{\partial \theta \partial \theta'} = \begin{bmatrix} \frac{\partial^2 C}{\partial \theta_1^2} & \cdots & \frac{\partial^2 C}{\partial \theta_1 \partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 C}{\partial \theta_p \partial \theta_1} & \cdots & \frac{\partial^2 C}{\partial \theta_p^2} \end{bmatrix} \quad (p \times p) \quad (3.3)$$

Definiendo  $\mathbf{q}(\theta) = \nabla C$  y  $\mathbf{Q}(\theta) = \nabla^2 C$  como notaciones funcionales.

En general, siempre es posible verificar si algún punto en el espacio de parámetros  $\Theta$  minimiza la función  $C$  localmente, examinando si las condiciones de primer y segundo orden se cumplen en ese punto. En general los métodos numéricos consisten de algoritmos de búsqueda implementados en computadores, los cuales tienden a encontrar una aproximación del mínimo, en un número finito de pasos o iteraciones. Existen diferentes algoritmos utilizados, pero la gran mayoría se basa en la aproximación cuadrática. Implementar un método de aproximación cuadrática requiere de dos elementos: la función objetivo  $C$  y el gradiente  $\mathbf{q} = \nabla C$  de la función, razón cual se denominan métodos del gradiente, el estimador  $\hat{\theta}$  es definido implícitamente por las condiciones de primer orden  $\mathbf{q}(\hat{\theta}) = 0$ , sujeta a las condiciones de segundo orden para un mínimo. Definamos entonces una aproximación cuadrática de  $C$  en el punto  $\theta^r \in \Theta$  por:

$$C_r(\theta) = \frac{1}{2} \theta' \mathbf{H}_r \theta + \mathbf{b}_r' \theta + c_r \quad (3.4)$$

donde  $\mathbf{H}_r$  ( $p \times p$ ) es una matriz simétrica definida positiva (elecciones alternativas de  $\mathbf{H}_r$  definirán distintos tipos de algoritmos), diferenciando 3.4 obtenemos:

$$\nabla C_r(\theta^r) = \mathbf{b}_r + \mathbf{H}_r \theta^r \quad (3.5)$$

Dado que  $\mathbf{H}_r$  y  $\mathbf{b}_r$  son elegidos de forma que  $C$  y  $C_r$  tengan el mismo gradiente en  $\theta^r$ , tenemos:

$$\nabla C_r(\theta^r) = \mathbf{q}(\theta^r)$$

escribiendo  $\mathbf{q}(\theta^r)$  de manera compacta como  $\mathbf{q}_r$ , obtenemos:

$$\mathbf{b}_r = \mathbf{q}_r - \mathbf{H}_r \theta^r \quad (3.6)$$

de manera que si se elige la función  $C_r$  tal que  $C(\theta^r) = C_r(\theta^r)$ , entonces  $C$  y  $C_r$  son tangentes en el punto  $\theta^r$ .

La idea radica básicamente en encontrar una aproximación del mínimo de  $C$  utilizando  $\theta^{*r}$ , el mínimo de  $C_r$  y  $\theta^{*r}$  es calculado igualando la ecuación 3.5 a cero y sustituyendo este valor en 3.6, se obtiene:

$$\theta^{*r} = -\mathbf{H}_r^{-1}\mathbf{b}_r = \theta^r - \mathbf{H}_r^{-1}\mathbf{q}_r \quad (3.7)$$

Tomando  $\theta^{*r}$  como el nuevo estimado del mínimo de  $C$ , sea  $\theta^{r+1} = \theta^{*r}$ , iterando la secuencia de cálculos desde  $r = 1$  y sucesivamente para  $r = 2, 3, \text{etc.}$ , es posible encontrar una aproximación del mínimo del criterio  $C$ .

Es necesario tomar en cuenta dos condiciones importantes: i) la selección del punto inicial  $\theta^1$  dentro del espacio de parámetros y ii) el criterio para detener la búsqueda, dado que cuando la secuencia tienda hacia el verdadero mínimo  $\hat{\theta}$ , el gradiente  $\mathbf{q}$  tenderá hacia  $\mathbf{q}(\hat{\theta}) = 0$ , de forma que la norma  $\|\theta^{r+1} - \theta^r\|$  tenderá hacia cero.

Lo que distingue a los distintos métodos es la elección de  $\mathbf{H}_r$ ; por ejemplo, la elección  $\mathbf{H}_r = \mathbf{I}$  deriva en el método de la escalera descendente, que es simple y robusto pero que converge lentamente dado que  $C_r$  es una pobre aproximación de  $C$ , otra elección de  $\mathbf{H}_r$  es  $\mathbf{Q}_r$  la matriz hesiana, que es utilizada en el algoritmo de Newton - Raphson (NR), ampliamente conocido. Sin embargo, este método adolece de dos problemas: i) La matriz hesiana es definida positiva en el mínimo pero no necesariamente en cualquier otro punto  $\theta \in \Theta$  y ii) en la mayoría de los casos la matriz hesiana no tiene una forma analítica.

Los algoritmos Quasi-Newton (QN) solucionan los dos problemas básicos del método NR, dado que utilizan una fórmula simple para actualizar  $\mathbf{H}_r^{-1}$  directamente, de una iteración a otra, de esta forma que no se requiere la matriz inversa, de forma explícita.

Sea la dirección de búsqueda en el paso  $r$  denotada por  $-\mathbf{B}_r\mathbf{q}_r$ , donde  $\mathbf{B}_r = \mathbf{H}_r^{-1}$ , utilizando el paso inicial más frecuente (la escalera descendente)  $\mathbf{B}_1 = \mathbf{I}$ , y definiendo  $\gamma_r = \mathbf{q}_{r+1} - \mathbf{q}_r$  y  $\delta_r = \theta^{r+1} - \theta^r = -\lambda_r\mathbf{B}_r\mathbf{q}_r$ , donde  $\delta_r$  representa la dirección de búsqueda en el paso  $r$  y  $\lambda_r \in \Re$  es un factor de escala que asegura que  $C(\theta^{r+1}) \leq C(\theta^r)$ . La clave de los métodos QN es la elección de  $\mathbf{B}_{r+1}$  que es definida positiva y satisface la ecuación de Quasi-Newton

$$\mathbf{B}_{r+1}\gamma_r = \delta_r \quad (3.8)$$

Para operativizar este método, es necesaria una fórmula de actualización para obtener  $\mathbf{B}_{r+1}$  dado  $\mathbf{B}_r$ . Existen diferentes esquemas, los más importantes pertenecen a la familia de algoritmos denominados Broyden, en esa formulación una matriz de rango 2 es agregada a  $\mathbf{B}_r$ , con la forma

$$\mathbf{B}_{r+1} = \mathbf{B}_r + \mathbf{u}_r (\mathbf{E}_r + \phi \mathbf{F}_r) \mathbf{u}'_r \quad (3.9)$$

donde

$$u_r = \begin{bmatrix} \delta_r & \mathbf{B}_r \gamma_r \end{bmatrix} (p \times 2) \quad (3.10)$$

$$\mathbf{E}_r = \begin{bmatrix} \frac{1}{\delta'_r \gamma_r} & 0 \\ 0 & -\frac{1}{\gamma'_r \mathbf{B}_r \gamma_r} \end{bmatrix}, \mathbf{F}_r = \begin{bmatrix} \frac{\gamma'_r \mathbf{B}_r \gamma_r}{(\delta'_r \gamma_r)^2} & -\frac{1}{\delta'_r \gamma_r} \\ -\frac{1}{\delta'_r \gamma_r} & -\frac{1}{\gamma'_r \mathbf{B}_r \gamma_r} \end{bmatrix}$$

son matrices  $2 \times 2$ , y  $\phi \geq 0$  es un escalar que define a los miembros de la familia. Nótese que  $\mathbf{F}_r$  es de rango 1 y que  $\mathbf{u}_r \mathbf{F}_r \mathbf{u}'_r = v_r v'_r$  donde:

$$v_r = \sqrt{\gamma'_r \mathbf{B}_r \gamma_r} \left( \frac{\delta_r}{\delta'_r \gamma_r} - \frac{\mathbf{B}_r \gamma_r}{\gamma'_r \mathbf{B}_r \gamma_r} \right) \quad (3.11)$$

Dado que  $v'_r \gamma_r = 0$  y

$$\mathbf{u}'_r \gamma_r = \begin{bmatrix} \delta'_r \gamma_r \\ \gamma'_r \mathbf{B}_r \gamma_r \end{bmatrix} \quad (3.12)$$

por lo tanto, la ecuación de Quasi - Newton se verifica para cualquier elección de  $\phi$ . Los miembros más importantes de la familia Broyden son el algoritmo Davidon- Fletcher- Powell (DFP), el cual toma  $\phi = 0$  y el Broyden- Goldfarb- Fletcher- Shanno (BFGS), para el cual  $\phi = 1$ . De ambos algoritmos, el BFGS se aplica mejor a la mayoría de los problemas de optimización de funciones no lineales y es ampliamente usado<sup>1</sup>.

## 3.2. Estimación de probabilidades utilizando el algoritmo GHK

Una de las desventajas en la estimación de modelos de elección discreta y en econometría en general, es el cálculo de integrales que no poseen una forma cerrada, como es el caso de la distribución normal, cuya integral no

---

<sup>1</sup>ver Broyden (1970)

posee una forma analítica, en estos casos, las integrales son estimadas por métodos numéricos y estocásticos. En el caso de modelos PMV, para efectuar las estimaciones, a partir de dos ecuaciones, se requiere la evaluación de integrales de orden superior o igual a dos. Actualmente, el cálculo de integrales de órdenes uno y dos no representan mayor dificultad, sin embargo, en el caso de la distribución normal multivariante, a partir del caso trivariante, el cálculo de las integrales respectivas implica un esfuerzo considerable en tiempo y número de pasos si se utilizan los métodos de aproximación polinómica, aproximación por cuadratura e integración de Monte Carlo. El método GHK desarrollado por Geweke, Hajivassilou y Keane entre 1989 y 1994 se muestra como un método exacto y sorprendentemente simple, el enfoque general se basa en la aproximación:

$$\Pr [a_1 < x_1 < b_1, a_2 < x_2 < b_2, \dots, a_J < x_J < b_J] \approx \frac{1}{R} \sum_{r=1}^R \prod_{j=1}^J Q_{rj} \quad (3.13)$$

donde  $R$  es el número de iteraciones,  $J$  es el orden de la probabilidad y  $Q_{rj}$  son probabilidades univariantes de fácil cálculo, estimadas a través de los siguientes pasos :

1. Factorización de la matriz de varianzas y covarianzas  $\Sigma$  de las variables aleatorias, usando la descomposición de Choleski  $\Sigma = LL'$ , donde  $L$  es una matriz triangular inferior.
2. Inicio del cálculo recursivo con

$$Q_{r1} = \Phi \left( \frac{b_1}{l_{11}} \right) - \Phi \left( \frac{a_1}{l_{11}} \right) \quad (3.14)$$

3. Generar un número aleatorio  $e_{r1}$  con distribución Normal truncada en el rango

$$\left[ A_{r1} = \frac{a_1}{l_{11}}, B_{r1} = \frac{b_1}{l_{11}} \right] \quad (3.15)$$

4. Para el resto de los casos:  $j = 2, \dots, J$ , se calcula:

$$A_{rj} = \frac{a_j - \sum_{m=1}^{j-1} l_{jm} e_{rm}}{l_{jj}} \quad (3.16)$$

$$B_{rj} = \frac{b_k - \sum_{m=1}^{j-1} l_{jm} e_{rm}}{l_{jj}} \quad (3.17)$$

entonces:

$$Q_{rj} = \Phi(B_{rk}) - \Phi(A_{rk})$$

En cada cálculo de nuevos límites, se debe generar un número aleatorio  $e_{rj}$  con distribución truncada en el rango  $[A_{rj}, B_{rj}]$ .

5. Se obtiene  $\prod_{j=1}^J Q_{rj}$  y se replica este proceso  $R$  veces. La probabilidad estimada es el promedio de las probabilidades simuladas y se recomienda el uso de 25 a 100 replicaciones.

El algoritmo GHK es impresionantemente exacto y rápido, fácil de programar y no necesita un elevado número de iteraciones para obtener una buena aproximación de la probabilidad buscada

### 3.3. Derivadas Numéricas

La mayoría de los métodos numéricos de optimización precisan de por lo menos las primeras derivadas para su aplicación y el método NR requiere, también, de las segundas derivadas. Las derivadas pueden calcularse analítica o numéricamente por diferenciación finita, claramente la disponibilidad de expresiones analíticas permite cálculos exactos y rápidos, sin embargo, el cálculo de derivadas analíticas es laborioso y, con frecuencia éstas son desconocidas, por lo que, en la práctica, las derivadas numéricas son más utilizadas.

Sea  $F(\cdot)$  una función de  $\mathbb{R}^p$  en  $\mathbb{R}$ , sea  $F(\theta^0)$  el valor disponible de  $F$  en  $\theta^0$ , entonces la aproximación de las primeras derivadas parciales en  $\theta^0$  está dada por:

$$\frac{\partial F(\theta^0)}{\partial \theta_i} \approx \frac{F(\theta_1^0, \dots, \theta_i^0 + \epsilon_i, \dots, \theta_p^0) - F(\theta_1^0, \dots, \theta_i^0, \dots, \theta_p^0)}{\epsilon_i} \quad (3.18)$$

si la evaluación se efectúa de forma asimétrica y si la evaluación se realiza simétricamente tenemos:

$$\frac{\partial F(\theta^0)}{\partial \theta_i} \approx \frac{F(\theta_1^0, \dots, \theta_i^0 + \epsilon_i, \dots, \theta_p^0) - F(\theta_1^0, \dots, \theta_i^0 - \epsilon_i, \dots, \theta_p^0)}{2\epsilon_i} \quad (3.19)$$

La derivada asimétrica requiere de  $p$  evaluaciones de la función  $F$  y en el caso de la derivada simétrica se requieren  $2p$  evaluaciones. En el caso de segundas derivadas numéricas, éstas son aproximadas por:

$$\frac{\partial F^2(\theta^0)}{\partial^2 \theta_i} \approx \frac{F(\theta_1^0, \dots, \theta_i^0 + \epsilon_i, \dots, \theta_p^0) - 2F(\theta_1^0, \dots, \theta_i^0, \dots, \theta_p^0) + F(\theta_1^0, \dots, \theta_i^0 - \epsilon_i, \dots, \theta_p^0)}{\epsilon_i^2} \quad (3.20)$$

para las derivadas directas y las segundas derivadas cruzadas en el caso asimétrico son:

$$\frac{\partial F^2(\theta^0)}{\partial \theta_i \partial \theta_j} \approx [F(\theta_1^0, \dots, \theta_i^0 + \epsilon_i, \dots, \theta_j^0 + \epsilon_j, \dots, \theta_p^0) - F(\theta_1^0, \dots, \theta_i^0 + \epsilon_i, \dots, \theta_p^0) - F(\theta_1^0, \dots, \theta_j^0 + \epsilon_j, \dots, \theta_p^0) + F(\theta_1^0, \dots, \theta_p^0)] / \epsilon_i \epsilon_j \quad (3.21)$$

y finalmente en el caso simétrico:

$$\frac{\partial F^2(\theta^0)}{\partial \theta_i \partial \theta_j} = [F(\theta_1^0, \dots, \theta_i^0 + \epsilon_i, \dots, \theta_j^0 + \epsilon_j, \dots, \theta_p^0) - F(\theta_1^0, \dots, \theta_i^0 - \epsilon_i, \dots, \theta_j^0 + \epsilon_j, \dots, \theta_p^0) - F(\theta_1^0, \dots, \theta_i^0 + \epsilon_i, \dots, \theta_j^0 - \epsilon_j, \dots, \theta_p^0) + F(\theta_1^0, \dots, \theta_i^0 - \epsilon_i, \dots, \theta_j^0 - \epsilon_j, \dots, \theta_p^0)] / 4\epsilon_i \epsilon_j \quad (3.22)$$

el uso de segundas derivadas numéricas implica la evaluación adicional de  $(3p^2 + p)/2$  y  $2p^2$  veces la función  $F(\cdot)$  en los casos asimétrico y simétrico respectivamente.

Finalmente  $\epsilon_i$  y  $\epsilon_j$  son obtenidos por la siguiente relación  $\epsilon_i = \max(\alpha |\theta_i|, \alpha)$ , donde  $\alpha = 10^{-5}$ .

# Capítulo 4

## Estimación del modelo PMV

### 4.1. Introducción

La estimación de modelos PMV ha demostrado ser inconsistente aplicando las diferentes versiones del método de variables instrumentales (véase Dagenais, 1999), por lo que estimaciones inconsistentes son encontradas si no se aplica el método de Máximo Verosimilitud con Información Completa (MVIC), única manera de evitar la inconsistencia. En este enfoque, todas las ecuaciones del sistema son estimadas de manera simultánea, incluyendo las correlaciones entre los residuos.

### 4.2. Primeras Derivadas Analíticas

La estimación de los modelos PMV es obtenida por el método de máximo verosimilitud. Si se utilizan los métodos comunes de gradiente (BFGS, DFP, etc.) son necesarias las primeras derivadas y en algunos casos como es el método de NR, se requieren las derivadas segundas. Las derivadas obtenidas de manera numérica requieren  $p$  evaluaciones adicionales del  $\log L$  (la función objetivo), en el caso asimétrico, donde

$$p = k + \frac{J(J-1)}{2} = \#\text{parámetros } \beta + \#\text{correlaciones } \omega_{lm}$$

es el número de parámetros desconocidos y  $2p$  evaluaciones en el caso simétrico, que es el más común. En el caso de las segundas derivadas, se requieren



$(3p^2 + p)/2$  evaluaciones en el caso asimétrico y  $2p^2$  evaluaciones en el caso simétrico, donde siguiendo la notación en la sección § , el logaritmo de verosimilitud del modelo PMV es:

$$\log L = \sum_{i=1}^n \log \Phi_J(\mathbf{w}_i, R_i) \quad (4.1)$$

*El mayor aporte de este trabajo es el de proporcionar formas analíticas de las primeras derivadas del  $\log L$  respecto a los parámetros en el modelo PMV y, por ende, mejorar la velocidad de estimación por MVIC.*

**Definición 4.1**

*Una variable aleatoria  $U$  posee distribución Normal  $J$ -variante  $N_J(\mu, \Sigma)$  si su función de densidad es:*

$$\phi_J(u | \mu, \Sigma) \equiv (2\pi)^{-J/2} |\Sigma|^{-1/2} e^{-1/2(u-\mu)'\Sigma^{-1}(u-\mu)}, u \in \mathfrak{R}^J \quad (4.2)$$

donde  $\mu \in \mathfrak{R}^J$  y  $\Sigma$  es una matriz  $J \times J$  definida positiva.

**Definición 4.2**

*Si  $U$  posee distribución Normal  $J$ -variante  $N_J(\mu, \Sigma)$ , entonces la variable  $U^* = H^{-1/2}(U - \mu)$ , donde los elementos de  $H$  se obtienen por  $h_{ii} = \sigma_{ii}$  y  $h_{ij} = 0$  si  $i \neq j$ , posee una distribución normal estándar  $J$ -variante  $N_J(0, \Omega)$ , en la que los elementos de la diagonal de  $\Omega$  son  $\omega_{ii} = 1$ . La función de densidad respectiva es:*

$$\phi_J(u^* | 0, \Omega) \equiv (2\pi)^{-J/2} |\Omega|^{-1/2} e^{-1/2(u^{*\prime}\Omega^{-1}u^*)}, u^* \in \mathfrak{R}^J \quad (4.3)$$

**Lema 4.1**

*Sea  $U$  un vector aleatorio  $J \times 1$  particionado en  $[U_1', U_2']'$ , donde  $U_1$  es  $m \times 1$  y  $U_2$  es  $(J - m) \times 1$ . Suponiendo que  $U \sim N_J(\mu, \Sigma)$ , donde*

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{bmatrix}$$

*están particionados de acuerdo a la partición de  $U$ .*

*Entonces las distribuciones marginales de  $U_1$  y  $U_2$  son :  $U_1 \sim N_m(\mu_1, \Sigma_{11})$  y  $U_2 \sim N_{J-m}(\mu_2, \Sigma_{22})$  respectivamente y la distribución condicional de  $U_2$ , dado  $U_1 = u_1$ , es  $U_2 | u_1 \sim N_{J-m}(\mu_{2|1}, \Sigma_{2|1})$ , donde:*

$$\mu_{2|1} \equiv E(U_2 | U_1 = u_1) = \mu_2 + \Sigma'_{12}\Sigma_{11}^{-1}(u_1 - \mu_1) \quad (4.4)$$

$$\Sigma_{2|1} = \Sigma_{22} - \Sigma'_{12}\Sigma_{11}^{-1}\Sigma_{12} \quad (4.5)$$

de manera análoga la distribución condicional de  $U_1$ , dado  $U_2 = u_2$ , es  $U_1 | u_2 \sim N_m(\mu_{1|2}, \Sigma_{1|2})$ , donde:

$$\mu_{1|2} \equiv E(U_1 | U_2 = u_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(u_2 - \mu_2) \quad (4.6)$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma'_{12} \quad (4.7)$$

*Demostración.* Poirier (1995) pag 122 □

El lema siguiente ha sido utilizado en Greene (2000, pg. 850), sin embargo, no se encontró una demostración explícita. En este trabajo proponemos una demostración directa que utiliza las propiedades de probabilidades condicionadas y las propiedades de la distribución normal multivariante.

**Lema 4.2**

Sea  $u$  un vector aleatorio bidimensional y supongamos que  $u \sim N_2(0, \Omega)$ , donde  $\omega_{ii} = 1$ , entonces

$$\frac{\partial \Phi_2(w_1, w_2, \sigma_{12})}{\partial \sigma_{12}} = \phi_2(w_1, w_2, \sigma_{12}) \quad (4.8)$$

*Demostración.* Por definición,

$$\Phi_2(w_1, w_2, \sigma_{12}) = \int_{-\infty}^{w_2} \int_{-\infty}^{w_1} \phi_2(t_1, t_2, \sigma_{12}) dt_1 dt_2 \quad (4.9)$$

dado que:

$$\phi_2(u_1, u_2, \sigma_{12}) = \phi_1(u_1) \phi_1(u_2 | \mu_{2|1}, \sigma_{2|1}) \quad (4.10)$$

donde  $\mu_{2|1}, \sigma_{2|1}$  son la media y varianza de la variable  $u_2$  condicionada a  $u_1 = t_1$  entonces:

$$\Phi_2(w_1, w_2, \sigma_{12}) = \int_{-\infty}^{w_2} \int_{-\infty}^{w_1} \phi_1(t_1) \phi_1(t_2 | \mu_{2|1}, \sigma_{2|1}) dt_1 dt_2 \quad (4.11)$$

$$\Phi_2(w_1, w_2, \sigma_{12}) = \int_{-\infty}^{w_1} \phi_1(t_1) \left[ \int_{-\infty}^{w_2} \phi_1(t_2 | \mu_{2|1}, \sigma_{2|1}) dt_2 \right] dt_1 \quad (4.12)$$

por el **Lema 4.1** y la **Definición 4.2** tenemos

$$\int_{-\infty}^{w_2} \phi_1(t_2 | \mu_{2|1}, \sigma_{2|1}) dt_2 = \int_{-\infty}^{w_2^*} \phi_1(t_2) dt_2 \quad (4.13)$$

donde

$$w_2^* = \frac{w_2 - \sigma_{12}t_1}{\sqrt{1 - \sigma_{12}^2}} \quad (4.14)$$

por lo tanto

$$\Phi_2(w_1, w_2, \sigma_{12}) = \int_{-\infty}^{w_1} \phi_1(t_1) \left[ \int_{-\infty}^{w_2^*} \phi_1(t_2) dt_2 \right] dt_1 \quad (4.15)$$

y derivando respecto a  $\sigma_{12}$  se obtiene

$$\frac{\partial \Phi_2(w_1, w_2, \sigma_{12})}{\partial \sigma_{12}} = \int_{-\infty}^{w_1} \phi_1(t_1) \phi_1(w_2^*) \frac{\partial w_2^*}{\partial \sigma_{12}} dt_1 \quad (4.16)$$

como

$$\frac{\partial w_2^*}{\partial \sigma_{12}} = (1 - \sigma_{12}^2)^{-1/2} \left( \frac{-t_1 + w_2 \sigma_{12}}{1 - \sigma_{12}^2} \right) \quad (4.17)$$

entonces

$$\begin{aligned} \frac{\partial \Phi_2(w_1, w_2, \sigma_{12})}{\partial \sigma_{12}} = \\ \int_{-\infty}^{w_1} (1 - \sigma_{12}^2)^{-1/2} \left( \frac{-t_1 + w_2 \sigma_{12}}{1 - \sigma_{12}^2} \right) \phi_1(t_1) \phi_1(w_2^*) dt_1 \end{aligned} \quad (4.18)$$

Por otra parte, dado que:

$$\frac{d\phi_1(t)}{dt} = -t\phi_1(t) \quad (4.19)$$

aplicando al integrando de la ecuación 4.16, tenemos

$$\frac{\partial [\phi_1(t_1) \phi_1(w_2^*)]}{\partial t_1} = -t_1 \phi_1(t_1) \phi_1(w_2^*) + w_2^* \frac{\sigma_{12}}{\sqrt{1 - \sigma_{12}^2}} \phi_1(t_1) \phi_1(w_2^*) \quad (4.20)$$

factorizando:

$$\frac{\partial [\phi_1(t_1) \phi_1(w_2^*)]}{\partial t_1} = \phi_1(t_1) \phi_1(w_2^*) \left[ -t_1 + w_2^* \frac{\sigma_{12}}{\sqrt{1 - \sigma_{12}^2}} \right] \quad (4.21)$$

$$\frac{\partial [\phi_1(t_1) \phi_1(w_2^*)]}{\partial t_1} = \phi_1(t_1) \phi_1(w_2^*) \left[ \frac{-t_1 + w_2 \sigma_{12}}{1 - \sigma_{12}^2} \right] \quad (4.22)$$

entonces

$$\int_{-\infty}^{w_1} \left( \frac{-t_1 + w_2 \sigma_{12}}{1 - \sigma_{12}^2} \right) \phi_1(t_1) \phi_1(w_2^*) dt_1 = \phi_1(t_1) \phi_1(w_2^*) \quad (4.23)$$

por lo tanto

$$(1 - \sigma_{12}^2)^{-1/2} \int_{-\infty}^{w_1} \left( \frac{-t_1 + w_2 \sigma_{12}}{1 - \sigma_{12}^2} \right) \phi_1(t_1) \phi_1(w_2^*) dt_1 = (1 - \sigma_{12}^2)^{-1/2} \phi_1(w_1) \phi_1(w_2^*) \quad (4.24)$$

Siendo que

$$\phi_1(w_2 | \mu_{2|1}, \sigma_{2|1}) = \phi_1(w_2^*) \frac{1}{\sqrt{1 - \sigma_{12}^2}} \quad (4.25)$$

obtenemos:

$$(1 - \sigma_{12}^2)^{-1/2} \int_{-\infty}^{w_1} \left( \frac{-t_1 + w_2 \sigma_{12}}{1 - \sigma_{12}^2} \right) \phi_1(t_1) \phi_1(w_2^*) dt_1 = \phi_1(w_1) \phi_1(w_2 | \mu_{2|1}, \sigma_{2|1}) \quad (4.26)$$

utilizando esto en 4.18 se obtiene:

$$\frac{\partial \Phi_2(w_1, w_2, \sigma_{12})}{\partial \sigma_{12}} = \phi_2(w_1, w_2, \sigma_{12}) \quad (4.27)$$

□

*Los siguientes teoremas son el aporte fundamental de este trabajo. No han sido demostrados anteriormente y tampoco existe una expresión analítica para las derivadas primeras del logaritmo de verosimilitud. Al demostrar estos resultados obtenemos un instrumento directo para el cálculo y estimación de los parámetros del modelo PMV.*

**Teorema 4.1**

Sea  $w_i^{j*} = (w_{i1}^{j*}, \dots, w_{i,j-1}^{j*}, w_{i,j+1}^{j*}, \dots, w_{iJ}^{j*})$ , el vector  $(w_{i1}, \dots, w_{i,j-1}, w_{i,j+1}, \dots, w_{iJ})$  estandarizado por la media y la matriz de varianzas y covarianzas de  $u_i^{j*} = (u_{i1}, \dots, u_{i,j-1}, u_{i,j+1}, \dots, u_{iJ})'$  condicional a  $u_{ij} = w_{ij}$ , y sea entonces  $R_i^{j*}$  la matriz de correlación del vector  $u_i^{j*}$  condicional a  $u_{ij} = w_{ij}$ . Entonces la derivada del  $\log L$  respecto  $\beta_j$  es

$$\frac{\partial \log L}{\partial \beta_j} = \sum_{i=1}^n \frac{\phi_1(w_{ij}) \Phi_{J-1}(\mathbf{w}_i^{j*}, R_i^{j*})}{\Phi_J(w_{i1}, \dots, w_{iJ}, R_i)} q_{ij} z'_{ij} \quad (4.28)$$

*Demostración.* Derivando el  $\log L$  obtenemos:

$$\frac{\partial \log L}{\partial \beta_j} = \sum_{i=1}^n \frac{1}{\Phi_J(\mathbf{w}_i, R_i)} \frac{\partial \Phi_J(\mathbf{w}_i, R_i)}{\partial w_{i1}} q_{ij} z'_{ij} \quad (4.29)$$

donde  $\Phi_J(\mathbf{w}_i, R_i)$  esta definido por:

$$\Phi_J(\mathbf{w}_i, R_i) = \int_{-\infty}^{w_{iJ}} \dots \int_{-\infty}^{w_{ij}} \dots \int_{-\infty}^{w_{i1}} \phi_J(t | 0, R_i) dt, t = (t_1, \dots, t_j, \dots, t_J) \in \mathfrak{R}^J \quad (4.30)$$

Para encontrar una forma analítica de la ecuación 4.29 en este caso, se requiere el uso de las propiedades de la probabilidad condicional, aplicadas al vector aleatorio  $\mathbf{u}_i$  reordenado de la siguiente forma

$\bar{\mathbf{u}}_i = [(u_{i1}, \dots, u_{i,j-1}, u_{i,j+1}, \dots, u_{iJ}), u_{ij}]'$ . Tenemos que  $\bar{\mathbf{u}}_i \sim N_J(0, \bar{R}_i)$  con una matriz de correlación  $\bar{R}_i$  particionada así:

$$\bar{R}_i = \begin{bmatrix} \bar{R}_{i11} & \bar{R}_{i12} \\ \bar{R}'_{i12} & 1 \end{bmatrix} \quad (4.31)$$

de acuerdo a la partición efectuada sobre  $\bar{\mathbf{u}}_i$ . Por su parte, se reduce el vector  $t = (t_1, \dots, t_j, \dots, t_J)$  en  $s = (t_1, \dots, t_{j-1}, t_{j+1}, \dots, t_J) \in \mathfrak{R}^{J-1}$  y se reescribe la función de densidad de  $\mathbf{u}_i$  en un producto de una función de densidad marginal y una condicional, entonces :

$$\Phi_J(\mathbf{w}_i, R_i) = \int_{-\infty}^{w_{ij}} \phi_1(t_j) \left[ \int_{-\infty}^{w_{i1}} \dots \int_{-\infty}^{w_{iJ}} \phi_{J-1}(s | \mu_i^j, \Sigma_i^j) ds \right] dt_j \quad (4.32)$$

donde  $\mu_i^j$  ( $\mu_{i|j}$ ) y  $\Sigma_i^j$  son la media y la matriz de varianzas y covarianzas del vector  $(u_{i1}, \dots, u_{i,j-1}, u_{i,j+1}, \dots, u_{iJ})'$  condicionado a  $u_{ij} = t_j$ . Aplicando el **lema 4.1** a  $\Sigma_i^j$ , obtenemos  $\Sigma_i^j = \bar{R}_{i11} - \bar{R}_{i12} \bar{R}'_{i12} = \{\sigma_i^j\}_{km}$   $k, m = 1, \dots, j-1, j+1, \dots, J$ , y la matriz de correlaciones respectiva es  $R_i^{j*} = \{r_i^{j*}\}_{km}$  donde:

$$r_{i,km}^{j*} = \frac{\sigma_{i,km}^j}{\sqrt{\sigma_{i,kk}^j \sigma_{i,mm}^j}} \quad (4.33)$$

y los límites

$$w_{id}^{j*} = \frac{w_{id} - \rho_{idj} w_{ij}}{\sqrt{\sigma_{idd}^j}} \quad d = 1, \dots, j-1, j+1, \dots, J$$

entonces

$$\frac{\partial \Phi_J(\mathbf{w}_i, R_i)}{\partial w_{ij}} = \phi_1(w_{ij}) \Phi_{J-1}(\mathbf{w}_i^{j*}, R_i^{j*}) \quad (4.34)$$

donde  $\mathbf{w}_i^{j*} = (w_{i1}^{j*}, \dots, w_{i,j-1}^{j*}, w_{i,j+1}^{j*}, \dots, w_{iJ}^{j*})$ . Finalmente, utilizando 4.34 en 4.29 obtenemos:

$$\frac{\partial \log L}{\partial \beta_j} = \sum_{i=1}^n \frac{\phi_1(w_{ij}) \Phi_{J-1}(\mathbf{w}_i^{j*}, R_i^{j*})}{\Phi_J(\mathbf{w}_i, R_i)} q_{ij} z'_{ij} \quad (4.35)$$

□

### Teorema 4.2

Sea  $w_i^{km*} = (w_{i1}^{km*}, \dots, w_{i,k-1}^{km*}, w_{i,k+1}^{km*}, \dots, w_{i,m-1}^{km*}, w_{i,m+1}^{km*}, \dots, w_{iJ}^{km*})$ , el vector  $(w_{i1}, \dots, w_{i,k-1}, w_{i,k+1}, \dots, w_{i,m-1}, w_{i,m+1}, \dots, w_{iJ})$  estandarizado por la media y la matriz de varianzas y covarianzas de  $u_i^{km*} = (u_{i1}, \dots, u_{i,k-1}, u_{i,k+1}, \dots, u_{i,m-1}, u_{i,m+1}, u_{iJ})'$  condicionado a  $(u_{ik}, u_{im})' = (w_{ik}, w_{im})'$ , y sea  $R_i^{km*}$  la matriz de correlación del vector  $u_i^{km*}$  condicionado a  $(u_{ik}, u_{im})' = (w_{ik}, w_{im})'$ . Entonces, la derivada del  $\log L$  respecto al coeficiente de correlación  $\omega_{km}$  es:

$$\frac{\partial \log L}{\partial \omega_{km}} = \sum_{i=1}^n \frac{\phi_2(w_{ik}, w_{im}, \rho_{i12}) \Phi_{J-2}(\mathbf{w}_i^{km*}, R_i^{km*})}{\Phi_J(\mathbf{w}_i, R_i)} q_{ik} q_{im} \quad (4.36)$$

*Demostración.* Derivando el  $\log L$  respecto a  $\omega_{km}$  obtenemos:

$$\frac{\partial \log L}{\partial \omega_{km}} = \sum_{i=1}^n \frac{1}{\Phi_J(\mathbf{w}_i, R_i)} \frac{\partial \Phi_J(\mathbf{w}_i, R_i)}{\partial \rho_{ikm}} q_{ik} q_{im} \quad (4.37)$$

donde  $\Phi_J(\mathbf{w}_i, R_i)$  es definido por:

$$\Phi_J(\mathbf{w}_i, R_i) = \int_{-\infty}^{w_{iJ}} \dots \int_{-\infty}^{w_{ik}} \dots \int_{-\infty}^{w_{im}} \dots \int_{-\infty}^{w_{i1}} \phi_J(t | 0, \Omega) dt \quad (4.38)$$

De manera similar a la demostración del teorema 1, se requiere reordenar el vector  $\mathbf{u}_i$ , aislando las variables correspondientes a los suíndices  $k$  y  $m$ . Así, suponiendo que  $k < m$  tenemos:

$\mathbf{u}_i = [(u_{i1}, \dots, u_{i,k-1}, u_{i,k+1}, \dots, u_{i,m-1}, u_{i,m+1}, u_{iJ}), u_{ik}, u_{im}]'$ , y  $\mathbf{u}_i \sim N_J(0, \mathbb{R}_i)$  con matriz de correlación definida por:

$$\mathbb{R}_i = \begin{bmatrix} \mathbb{R}_{i11} & & & \mathbb{R}_{i12} \\ & \left[ \begin{array}{cc} 1 & \rho_{ikm} \\ \rho_{ikm} & 1 \end{array} \right] & & \\ & & & \end{bmatrix} \quad (4.39)$$

particionada de acuerdo a la partición de  $\mathbf{u}_i$ , siendo que  $t = (t_1, \dots, t_j, \dots, t_J) \in \mathfrak{R}^J$ , sea  $v = (t_1, \dots, t_{k-1}, t_{k+1}, \dots, t_{m-1}, t_{m+1}, \dots, t_J) \in \mathfrak{R}^{J-2}$ , entonces por propiedad de condicional y normal  $\Phi_J(\mathbf{w}_i, R_i)$  puede escribirse como:

$$\Phi_J(\mathbf{w}_i, R_i) = \int_{-\infty}^{w_{iJ}} \cdots \int_{-\infty}^{w_{i1}} \phi_{J-2}(v | 0, \mathbb{R}_{i11}) \left[ \int_{-\infty}^{w_{im}} \int_{-\infty}^{w_{ik}} \phi_2(t_m, t_k | M_i^{km}, \Sigma_i^{km}) dt_m dt_k \right] dv \quad (4.40)$$

donde  $M_i^{km}$  y  $\Sigma_i^{km}$  son la media y matriz de varianzas y covarianzas de  $(u_{ik}, u_{im})'$  condicionada a  $(u_{i1}, \dots, u_{i,k-1}, u_{i,k+1}, \dots, u_{i,m-1}, u_{i,m+1}, u_{iJ}) = v$ , y siendo que :

$$\int_{-\infty}^{w_{im}} \int_{-\infty}^{w_{ik}} \phi_2(t_m, t_k | \mu_i^{km}, \Sigma_i^{km}) dt_m dt_k = \int_{-\infty}^{w_{im}^*} \int_{-\infty}^{w_{ik}^*} \phi_2(t_m, t_k | 0, \rho_{ikm}^*) dt_m dt_k \quad (4.41)$$

donde  $\rho_{ikm}^*$  es el coeficiente de correlación entre  $u_{ik}, u_{im}$  condicional a  $(u_{i1}, \dots, u_{i,k-1}, u_{i,k+1}, \dots, u_{i,m-1}, u_{i,m+1}, \dots, u_{iJ})' = v$ , definido por el lema 4.2 entonces  $M_i^{km} = \mathbf{R}'_{i12} \mathbf{R}_{i11}^{-1} v = \{\mu_i^{km}\}'_c$  con  $c = 1, 2$  y  $\Sigma_i^{km} = \begin{bmatrix} 1 & \rho_{ikm} \\ \rho_{ikm} & 1 \end{bmatrix} - \mathbf{R}'_{i12} \mathbf{R}_{i11}^{-1} \mathbf{R}_{i12} = \{\sigma_i^{km}\}_{ab}$  con  $a, b = 1, 2$ . El coeficiente de correlación respectivo es:

$$\rho_{ikm}^* = \frac{\sigma_{i12}^{km}}{\sqrt{\sigma_{i11}^{km} \sigma_{i22}^{km}}} \quad (4.42)$$

y los límites son:

$$w_{ik}^* = \frac{w_{ik} - \mu_{i1}^{km}}{\sqrt{\sigma_{i11}^{km}}} \quad (4.43)$$

$$w_{im}^* = \frac{w_{im} - \mu_{i2}^{km}}{\sqrt{\sigma_{i22}^{km}}} \quad (4.44)$$

dado el lema 4.2

$$\frac{\partial \Phi_2(w_{ik}^*, w_{im}^*, \rho_{ikm}^*)}{\partial \rho_{ikm}^*} = \phi_2(w_{ik}^*, w_{im}^*, \rho_{ikm}^*) \quad (4.45)$$

y dado que

$$\frac{\partial \rho_{ikm}^*}{\partial \rho_{ikm}} = \frac{1}{\sqrt{\sigma_{i11}^{km} \sigma_{i22}^{km}}} \quad (4.46)$$

aplicando esto a la derivada de 4.41, obtenemos:

$$\begin{aligned} \frac{\partial \int_{-\infty}^{wim} \int_{-\infty}^{wik} \phi_2(t_m, t_k \mid \mu_i^{km}, \Sigma_i^{km}) dt_m dt_k}{\partial \rho_{ikm}} &= \\ \frac{\phi_2(w_{ik}^*, w_{im}^*, \rho_{ikm}^*)}{\sqrt{\sigma_{i11}^{km} \sigma_{i22}^{km}}} &= \\ \phi_2(w_{im}, w_{ik} \mid M_i^{km}, \Sigma_i^{km}) & \end{aligned}$$

entonces la derivada de  $\Phi_J(\mathbf{w}_i, R_i)$  respecto de  $\rho_{ikm}$  es:

$$\frac{\partial \Phi_J(\mathbf{w}_i, R_i)}{\partial \rho_{ikm}} = \int_{-\infty}^{wiJ} \dots \int_{-\infty}^{wi1} \phi_{J-2}(v) \phi_2(w_{ik}, w_{im} \mid M_i^{km}, \Sigma_i^{km}) dv \quad (4.47)$$

$$\frac{\partial \Phi_J(\mathbf{w}_i, R_i)}{\partial \rho_{ikm}} = \int_{-\infty}^{wiJ} \dots \int_{-\infty}^{wi1} \phi_J(t \mid 0, R_i) dv \quad (4.48)$$

y utilizando propiedades de la normal condicionada obtenemos:

$$\begin{aligned} \frac{\partial \Phi_J(\mathbf{w}_i, R_i)}{\partial \rho_{ikm}} &= \\ \int_{-\infty}^{wiJ} \dots \int_{-\infty}^{wi1} \phi_2(w_{ik}, w_{im}, \rho_{ikm}) \phi_{J-2}(w_{ik}, w_{im} \mid M_i^{km*}, \Sigma_i^{km*}) dv & \end{aligned} \quad (4.49)$$

$$\frac{\partial \Phi_J(\mathbf{w}_i, R_i)}{\partial \rho_{ikm}} = \phi_2(w_{ik}, w_{im}, \rho_{ikm}) \Phi_{J-2}(\mathbf{w}_i^{km*}, R_i^{km*}) \quad (4.50)$$

donde  $\mathbf{w}_i^{km*} = (w_{i1}^{km*}, \dots, w_{i,k-1}^{km*}, w_{i,k+1}^{km*}, \dots, w_{i,m-1}^{km*}, w_{i,m+1}^{km*}, \dots, w_{iJ}^{km*})$ , y definiendo

$\Sigma_i^{km} = \{\sigma_{iab}^{km}\}$  como  $\mathbf{R}_{i11} - \mathbf{R}_{i12} \begin{bmatrix} 1 & \rho_{ikm} \\ \rho_{ikm} & 1 \end{bmatrix}^{-1} \mathbf{R}'_{i12}$  obtenemos  $R_i^{km*} = \{r_{iab}^{km*}\}$  por la relación:

$$r_{iab}^{km*} = \frac{\sigma_{iab}^{km}}{\sqrt{\sigma_{iab}^j \sigma_{iab}^j}} \quad (4.51)$$



y los límites son:

$$w_{id}^{j*} = \frac{w_{id} - \frac{\rho_{idk}(w_{ik} - \rho_{ikm}w_{im}) + \rho_{idm}(w_{im} - \rho_{ikm}w_{ik})}{1 - \rho_{km}^2}}{\sqrt{\sigma_{idd}^{km}}} \quad (4.52)$$

finalmente, reemplazando el resultado 4.50 en 4.37 obtenemos:

$$\frac{\partial \log L}{\partial \omega_{km}} = \sum_{i=1}^n \frac{\phi_2(w_{ik}, w_{im}, \rho_{i12}) \Phi_{J-2}(\mathbf{w}_i^{km*}, R_i^{km*})}{\Phi_J(\mathbf{w}_i, R_i)} q_{ik} q_{im} \quad (4.53)$$

□

Habiéndose obtenido expresiones analíticas para el logaritmo de verosimilitud, y para las primeras derivadas del mismo es posible implementar un algoritmo como BFGS para encontrar los estimadores MVIC, para ilustrar este procedimiento la sección §5 describe el análisis de la pobreza y vivienda utilizando un modelo PMV de orden tres.

### Ejemplos

Habiéndose demostrado el **teorema 4.1** y el **teorema 4.2** existen las herramientas necesarias para encontrar las primeras derivadas del Logaritmo de verosimilitud para cualquier PMV. La derivada parcial del  $\log L$  respecto a los parámetros de la ecuación  $j = 1, 2, 3$  en el caso triprobit es:

$$\frac{\partial \log L}{\partial \beta_j} = \sum_{i=1}^n \frac{q_{ij} g_{ij}}{\Phi_3} z'_{ij}, \quad j = 1, 2, 3 \quad (4.54)$$

donde

$$\begin{aligned} g_{i1} &= \phi(w_{i1}) \Phi_2 \left( \frac{w_{i2} - \rho_{i12}w_{i1}}{\sqrt{1 - \rho_{i12}^2}}, \frac{w_{i3} - \rho_{i13}w_{i1}}{\sqrt{1 - \rho_{i13}^2}}, \frac{\rho_{i23} - \rho_{i12}\rho_{i13}}{\sqrt{(1 - \rho_{i12}^2)(1 - \rho_{i13}^2)}} \right) \\ g_{i2} &= \phi(w_{i2}) \Phi_2 \left( \frac{w_{i1} - \rho_{i12}w_{i2}}{\sqrt{1 - \rho_{i12}^2}}, \frac{w_{i3} - \rho_{i23}w_{i2}}{\sqrt{1 - \rho_{i23}^2}}, \frac{\rho_{i13} - \rho_{i12}\rho_{i23}}{\sqrt{(1 - \rho_{i12}^2)(1 - \rho_{i23}^2)}} \right) \\ g_{i3} &= \phi(w_{i3}) \Phi_2 \left( \frac{w_{i1} - \rho_{i13}w_{i3}}{\sqrt{1 - \rho_{i13}^2}}, \frac{w_{i2} - \rho_{i13}w_{i3}}{\sqrt{1 - \rho_{i13}^2}}, \frac{\rho_{i12} - \rho_{i13}\rho_{i23}}{\sqrt{(1 - \rho_{i13}^2)(1 - \rho_{i23}^2)}} \right) \end{aligned}$$

En cuanto a los coeficientes de correlación la derivada de  $\log L$  respecto del coeficiente de correlación  $\omega_{km}$  es :

$$\frac{\partial \log L}{\partial \omega_{km}} = \sum_{i=1}^n \frac{\phi_2(w_{ik}, w_{im}, \rho_{ikm}) \Phi_1(w_i^{km*})}{\Phi_{i3}} q_{ik} q_{im} \quad (4.55)$$

donde

$$\begin{aligned}
w_i^{12*} &= \frac{(1 - \rho_{i12}^2) w_{i3} - (\rho_{i13} - \rho_{i23}\rho_{i12}) w_{i1} - (\rho_{i23} - \rho_{i13}\rho_{i12}) w_{i2}}{\sqrt{1 - \rho_{i12}^2} \sqrt{1 - \rho_{i12}^2 - \rho_{i13}^2 - \rho_{i23}^2 + 2\rho_{i12}\rho_{i13}\rho_{i23}}} \\
w_i^{13*} &= \frac{(1 - \rho_{i13}^2) w_{i2} - (\rho_{i12} - \rho_{i23}\rho_{i13}) w_{i1} - (\rho_{i23} - \rho_{i13}\rho_{i12}) w_{i3}}{\sqrt{1 - \rho_{i13}^2} \sqrt{1 - \rho_{i12}^2 - \rho_{i13}^2 - \rho_{i23}^2 + 2\rho_{i12}\rho_{i13}\rho_{i23}}} \\
w_i^{23*} &= \frac{(1 - \rho_{i23}^2) w_{i1} - (\rho_{i12} - \rho_{i23}\rho_{i13}) w_{i2} - (\rho_{i13} - \rho_{i23}\rho_{i12}) w_{i3}}{\sqrt{1 - \rho_{i23}^2} \sqrt{1 - \rho_{i12}^2 - \rho_{i13}^2 - \rho_{i23}^2 + 2\rho_{i12}\rho_{i13}\rho_{i23}}}
\end{aligned}$$

### 4.3. Efectos Marginales

Una vez estimados los parámetros del modelo, es posible obtener, al igual que en los modelos probit univariante, los efectos marginales de los distintos tipos de probabilidades<sup>1</sup>. A diferencia de los modelos probit univariantes, los modelos PMV de grado mayor o igual a dos poseen una infinidad de posibilidades de efectos marginales que pueden clasificarse en tres tipos básicos: efectos marginales: incondicionales, conjuntos y condicionales.

#### Efectos Marginales Incondicionales

Las distribuciones incondicionales están dadas, en el caso de los modelos PMV, por distribuciones normales univariantes, por lo tanto tenemos:

$$E[y_j | z_j] = \Pr[y_j = 1 | z_j] = \Phi(z_j \beta_j) \quad (4.56)$$

y los efectos marginales sobre estas probabilidades  $EMI = \{EMI\}_{mj}$ ,  $m = 1, \dots, k_j$ ,  $j = 1, \dots, J$ , pueden ser obtenidos, en el caso continuo, de la manera usual con la formula:

$$EMI_j = \frac{\partial \Phi(z_j \beta_j)}{\partial z_j} = \phi(z_j) \beta_j, \quad j = 1, \dots, J \quad (4.57)$$

donde  $EMI_j$  es un vector columna  $k_j \times 1$ .

Estos efectos, en el caso discreto, se obtienen mediante el ratio del incremento en la probabilidad sobre el incremento en covariada  $z_{jm}$  (la covariada  $m$  de la ecuación  $j$ ), es decir:

$$EMI_{mj} \approx \frac{\Delta \Phi(z_j \beta_j)}{\Delta z_{mj}} = \Phi(\mathbf{z}_j^1 \beta_j) - \Phi(\mathbf{z}_j^0 \beta_j), \quad j = 1, \dots, J \quad m = 1, \dots, k_j \quad (4.58)$$

---

<sup>1</sup>Por motivos de orden y simplicidad en esta sección se omiten los subíndices correspondientes a las observaciones.

donde  $\mathbf{z}_j^0 = (z_{1j}, \dots, z_{m-1,j}, 0, z_{m+1,j}, \dots, z_{k_j j})$  y  $\mathbf{z}_j^1 = (z_{1j}, \dots, z_{m-1,j}, 1, z_{m+1,j}, \dots, z_{k_j j})$

En el caso PMV existen  $k = \sum_{j=1}^J k_j$  efectos marginales de este tipo, donde  $k_j$  es el número de covariadas de la ecuación  $j$ .

### Efectos Marginales Conjuntos

Los efectos marginales conjuntos son obtenidos considerando la distribución conjunta de los resultados del modelo PMV, es decir, tomando en cuenta la probabilidad conjunta de los distintos resultados  $\Pr(y_1, \dots, y_J) = \Phi(w_1, \dots, w_J, R)$ , definamos  $\zeta = [z_1 \cup z_2 \cup \dots \cup z_J] = \{\zeta\}'_{\kappa}$ ,  $\kappa = 1, \dots, K$  y  $K$  es el número total de covariadas distintas, por lo que  $\zeta_{\kappa}$  es una covariada que puede repetirse en una o más ecuaciones. Los efectos marginales consecuencia directa del resultado 4.34, para el caso continuo son:

$$EMC_{\kappa} = \frac{\partial \Phi(w_1, \dots, w_J, R)}{\partial \zeta_{\kappa}} = \sum_{j=1}^J \gamma_j \beta_{\kappa j} \phi_1(w_j) \Phi_{J-1}(\mathbf{w}^{j*}, R^{j*}) \quad \kappa = 1, \dots, K \quad (4.59)$$

donde  $\gamma_j$  es uno si la covariada  $\zeta_{\kappa}$  se encuentra en la ecuación  $j$  y cero en otro caso, y  $\beta_{\kappa j}$  es el coeficiente de la covariada  $\zeta_{\kappa}$  en la ecuación  $j$ . En el caso discreto la aproximación está dada por:

$$\frac{\Delta \Phi(w_1, \dots, w_J, R)}{\Delta \zeta_{\kappa}} = \frac{\Phi_J(q_1 z_1^1 \beta_1, \dots, q_J z_J^1 \beta_J, R) - \Phi_J(q_1 z_1^0 \beta_1, \dots, q_J z_J^0 \beta_J, R)}{\Delta \zeta_{\kappa}} \quad \kappa = 1, \dots, K$$

donde  $z_j = (z_{1j}, \dots, z_{m-1,j}, z_{mj}, z_{m+1,j}, \dots, z_{k_j j})$  con  $z_{mj} = \zeta_{\kappa}$  y  $m$  entre 1 y  $k_j$ , de esta manera  $z_j^1 = (z_{1j}, \dots, z_{m-1,j}, 1, z_{m+1,j}, \dots, z_{k_j j})$  y  $z_j^0 = (z_{1j}, \dots, z_{m-1,j}, 0, z_{m+1,j}, \dots, z_{k_j j})$ . Siendo que cada  $y_j$  asume dos valores, existen  $2^J K$  efectos marginales conjuntos.

### Efectos Marginales Condicionales

La tercera clase de efectos marginales utilizados son los gradientes de las probabilidades condicionales o efectos marginales condicionales, que son

obtenidos de la probabilidad:

$$\Pr(y_j = 1 | \mathbf{y}^j, \mathbf{z}) = \frac{\Pr(y_j = 1, \mathbf{y}^j | \mathbf{z})}{\Pr(\mathbf{y}^j | \mathbf{z})} \quad (4.60)$$

donde  $\mathbf{y}^j = (y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_J)$ . Escribiendo esta expresión en términos de la normal obtenemos:

$$\Pr(y_j = 1 | \mathbf{y}^j, \mathbf{z}) = \frac{\Phi(\mathbf{w}, R)}{\Phi(w_1, \dots, w_{j-1}, w_{j+1}, \dots, w_J, R^j)} \quad (4.61)$$

donde  $q_j = 1$ . Esta expresión aparentemente peligrosa y difícil de derivar, se puede reescribir como una normal acumulada de la siguiente manera, aplicando la propiedad de probabilidades condicionales:

$$\Pr(y_j = 1 | \mathbf{y}^j, \mathbf{z}) = \Phi(w_j^*) = \Phi[w_j^*(\zeta)] \quad (4.62)$$

donde  $w_j^*$  es  $w_j$  estandarizado por la media y la varianza de  $u_j$  condicionada a los restantes términos de error. Finalmente, derivando la última expresión respecto a  $\zeta_\kappa$ , el  $EMC_\kappa$  es:

$$EMC_\kappa = \frac{\partial \Pr(y_j = 1 | \mathbf{y}^j, \mathbf{z})}{\partial \zeta_\kappa} = \phi(w_j^*(\zeta)) \frac{\partial w_j^*(\zeta)}{\partial \zeta_\kappa} \quad (4.63)$$

y de manera análoga, en el caso discreto, puede ser obtenido por:

$$EMC_\kappa = \frac{\partial \Pr(y_j = 1 | \mathbf{y}^j, \mathbf{z})}{\partial \zeta_\kappa} = \Phi(w_j^*(\zeta^1)) - \Phi(w_j^*(\zeta^0)) \quad (4.64)$$

donde  $\zeta^1 = (\zeta_1, \dots, \zeta_{\kappa-1}, 1, \zeta_{\kappa+1}, \dots, \zeta_K)$  y  $\zeta^0 = (\zeta_1, \dots, \zeta_{\kappa-1}, 0, \zeta_{\kappa+1}, \dots, \zeta_K)$  y en este caso existen  $2^{J-1}JK$  distintos efectos marginales condicionales.

### Matriz de Covarianzas de los Efectos Marginales

La expresión analítica de cada coeficiente de la matriz de covarianzas es compleja de escribir, contiene varios términos que la tornan de gran tamaño y dificultad. Sin embargo, es innecesario calcularlas de forma explícita ya que los efectos marginales se calculan solo una vez y para una sola observación. Por ello, éstas son en general halladas de manera numérica y utilizando el método Delta<sup>2</sup>.

$$Var(EM) = G \times Var(\hat{\theta}) \times G' \quad (4.65)$$

---

<sup>2</sup>vease Greene (2000)

donde  $EM$  es el efecto marginal de dimensión  $e \times 1$ ,  $\hat{\theta}$  es el conjunto de estimados MVIC incluyendo correlaciones y  $G = \frac{\partial EM}{\partial \theta'}$  es la matriz jacobiana de los efectos marginales evaluada en  $\hat{\theta}$ .

# Capítulo 5

## Aplicación

En los programas informáticos especializados en modelos probit multivariante de grado superior a dos, los algoritmos más frecuentemente para encontrar los estimadores MVIC son los de gradiente. Como ya se explicó en el III capítulo, estos procedimientos precisan calcular numéricamente las derivadas primeras, en los casos Broyden y las derivadas segundas en el caso NR. Esto implica un número elevado de cálculos, que se incrementan exponencialmente con el número de ecuaciones.

Por ejemplo, en los métodos Broyden, con una base de datos de 1000 observaciones, es necesario evaluar 1,000 probabilidades para obtener un valor en el  $\log L$ , además, es necesario el cálculo numérico de primeras derivadas para cada parámetro, por lo tanto, el uso de derivadas numéricas en un problema de 4 parámetros, el modelo tetraprobit contiene 10 parámetros (4 parámetros ecuacionales y 6 correlaciones independientes), esto implica 21,000 evaluaciones para obtener las primeras derivadas, incluyendo la evaluación del  $\log L$ . Este proceso se repite varias veces, hasta alcanzar convergencia, si son necesarias 20 iteraciones, entonces se requieren 420,000 evaluaciones de las integrales. El tiempo promedio que necesita un ordenador estándar (Pentium II) para calcular cada probabilidad es de 0.01 segundos aproximadamente, el tiempo necesario para estimar los parámetros es 1,1667h. (1h,10 mín), si la base tiene el doble de datos, el tiempo se dobla. El uso de este proceso, en bases de datos de carácter social como son las encuestas de hogares (MECOVI), puede representar una cantidad enorme de tiempo.

Utilizando los resultados presentados en este trabajo, a saber, las derivadas analíticas de la función  $\log L$ , se requieren las 1,000 evaluaciones del logaritmo de verosimilitud, a la velocidad de 0.01 segundos cada una. En cambio,

para el cálculo de las derivadas analíticas, tan sólo son necesarias 10,000 evaluaciones de integrales, con un orden menor, en este caso tres (al encontrar la derivada analítica, el orden de la integral se reduce en uno) y , por lo tanto, el tiempo que requiere el cálculo de las 10000 integrales de orden 3 es la mitad del requerido por las integrales de orden 4 (vale decir 0.005 segundos cada una ). Entonces, para cada iteración es necesario un tiempo total de 60 segundos y 20 iteraciones necesitarían 20 minutos. Sin embargo, este periodo se reduce debido a que, el tiempo de evaluación de las probabilidades de las primeras derivadas puede ser aún inferior.

## 5.1. La relación entre salud y pobreza: una aplicación

Los modelos PMV tienen una gran cantidad de aplicaciones, no solo en economía sino también en psicología, medicina, química, física y, en general, en cualquier área de estudio que precise el análisis de variables discretas.

Las dos áreas principales de la economía que utilizan, ampliamente, los modelos PMV son: i) estudios de microeconomía, a saber, demandas de bienes durables, comportamiento del consumidor sobre selecciones complementarias, acceso a servicios de salud, vivienda y pobreza, etc. y ii) estudios de macroeconomía, como son los estudios de las relaciones causales entre crisis financiera, bancaria, cambiaria y otras.

La pobreza es una variable socioeconómica observada, en la mayoría de los casos, como variable índice, se la mide por línea de pobreza o pobreza medida por NBI. De manera análoga, los indicadores de salud son variables discretas, dado que es imposible medir el grado de dolor e insatisfacción de las personas de forma continua y las personas están o no enfermas. Analizar la relación entre variables de salud y de pobreza es un caso típico de aplicación de modelos PMV.

*El análisis Probit es una técnica comúnmente usada para estudiar el efecto de estímulos (pobreza), sobre la probabilidad de respuesta (enfermedad), sin embargo en muchas aplicaciones, el estímulo podría afectar a más de un sistema y se podría entonces apreciar la respuesta de cada sistema de manera separada, por ejemplo, nosotros esperaríamos que la pobreza afectará de manera directa no solo a la enfermedad de un solo miembro de la familia sino que existe la posibilidad de contagio y por lo tanto la mejor forma es modelar*

la correlaciones de estas variables , es tambien importante no confundir una situación multivariante con una situación policotómic, en la cual dos o más grados son de hecho solo respuestas de un sistema unico a diferentes niveles (por ejemplo en enfermedades saludable, enfermo, y muerto.

Para aplicar e ilustrar la mejora introducida en la estimación del modelo PMV, en esta sección se analiza el impacto de la pobreza, variable de corto plazo, sobre tres diferentes enfermedades coyunturales, el estudio fue implementado con información del área urbana y rural de Bolivia, presentada en la encuesta de hogares 2001 del Instituto Nacional de Estadística (INE). Definamos *eda* la variable dicotómica que asume el valor **1** si se presentaron Enfermedades Diarréicas Agudas (EDA) en algun niño menor a 5 años del hogar y **0** en otro caso (eoc), la variable dicotómica *era* que asume el valor **1** si se presentaron Enfermedades Respiratorias Agudas (ERAS) en algun niño menor a 5 años del hogar y **0** eoc, *otras* la variable dicotómica que asume el valor **1** si se presentaron Enfermedades Coyunturales en otros miembros del hogar y **0** eoc y la variable dicotómica de *pobre* que toma el valor **1** si el hogar es catalogado pobre por el nivel de ingreso, según la clasificación del INE en la MECOVI 2001 y **0** eoc. La siguiente base de datos fue extraída al procesar la encuesta:



## Base de Datos

pobre	eda	era	otras	frec.	pobre	eda	era	otras	frec.
0	0	0	0	1166	1	0	0	0	1286
0	0	0	1	580	1	0	0	1	765
0	0	1	0	141	1	0	1	0	252
0	0	1	1	81	1	0	1	1	207
0	1	0	0	50	1	1	0	0	138
0	1	0	1	23	1	1	0	1	99
0	1	1	0	53	1	1	1	0	220
0	1	1	1	55	1	1	1	1	217

Como covariada se utilizó *pobre*, que afecta a las tres variables endógenas (*eda*, *era*, *otras*), ninguna variable endógena puede ser considerada razonablemente covariada, debido a que no existe una justificación económica ni social para hacerlo. En efecto, las variables bajo análisis, son observaciones realizadas en un periodo suficientemente corto, que impide el desarrollo de causalidad entre ellas y el impacto de la pobreza, sobre todas, es evidente.

Tomando en cuenta estos supuestos, el modelo en términos de variables no observables es :

$$\begin{aligned}
 eda_i^* &= \beta_{11} + \beta_{12}pobre_i + u_{i1} \\
 era_i^* &= \beta_{21} + \beta_{22}pobre_i + u_{i2} \\
 otras_i^* &= \beta_{31} + \beta_{32}pobre_i + u_{i3}
 \end{aligned}
 \tag{5.1}$$

Para obtener las estimaciones MVIC se desarrolló un programa dentro del paquete informático Ox Profesional versión 3.2 y su editor Oxedit 1.61<sup>1</sup>. La estimación MVIC del modelo triprobit se realizó según los siguientes pasos:

1. Se definió la función a maximizar  $\log L$ , de acuerdo a la ecuación 4.1.

---

<sup>1</sup>Ambos desarrollados por el Profesor Jurgen Doornick. Ox posee un lenguaje matricial orientado a objetos con una extensa librería de funciones matemáticas y estadísticas y posee algunas ventajas respecto a la velocidad sobre la mayoría de los paquetes similares. Su sintaxis funcional es similar a la del GAUSS. Ver en <http://www.oxedit.com>, y <http://www.nuff.ox.ac.uk/Users/Doornick/index.html>.

2. Los valores iniciales necesarios para el inicio del procedimiento de búsqueda fueron obtenidos transformando las estimaciones uniecuacionales MCO (Mínimos Cuadrados Ordinarios) por la aproximación a las estimaciones probit univariante:  $2,5\beta_{MCO} - 1,5$  para los coeficientes de los términos constantes y  $2,5\beta_{MCO}$  para los demás coeficientes, formando el vector de valores iniciales  $\theta^1$ .
3. Utilizando estos valores iniciales y la expresión definida para  $\log L$  se aplicó el algoritmo BFGS, que está implementado en Ox, como el procedimiento *MAXBFGS()* para obtener los estimadores MVIC. Para el cálculo de las probabilidades trivariantes, se desarrolló el algoritmo GHK, en el programa.
4. Utilizando la notación de la sección §3.1 y las definiciones siguientes: sea: el gradiente de  $\log L$   $\mathbf{q}_r(\theta) = \{q_r\}_j$ ,  $\theta_r = \{\theta_r\}_j$ , y  $\zeta_r(\theta) = \mathbf{B}_r \mathbf{q}_r = \{\zeta_r\}_j$  donde  $j = 1, \dots, p$ , este último define la dirección y magnitud del cambio. Los criterios de convergencia usados por Ox son los siguientes: el primero basado en las elasticidades  $\left( \frac{\partial \log L}{\partial \log |\theta_{rj}|} \right)$

$$\begin{aligned} |q_{r,j}\theta_{rj}| &\leq \varepsilon, \text{ para todo } j \text{ con } \theta_{ij} \neq 0 \\ |q_{i,j}| &\leq \varepsilon, \text{ para todo } j \text{ con } \theta_{ij} = 0 \end{aligned} \quad (5.2)$$

y el segundo basado en el cambio relativo en los valores de los parámetros

$$\begin{aligned} |\zeta_{r+1,j}| &\leq 10 |\theta_{r,j}| \varepsilon, \text{ para todo } j \text{ con } \theta_{ij} \neq 0 \\ |\zeta_{r+1,j}| &\leq 10\varepsilon, \text{ para todo } j \text{ con } \theta_{ij} = 0 \end{aligned} \quad (5.3)$$

si ambos criterios son obtenidos para  $\varepsilon = \varepsilon_1$  los estimadores han convergido de manera fuerte a los verdaderos parámetros. En otro caso, si el paso  $\lambda_r \leq 7,105 \times 10^{-15}$  y el primer criterio es obtenido para  $\varepsilon = \varepsilon_2$ , entonces los estimadores han convergido débilmente a los verdaderos parámetros y se utilizó  $\varepsilon_1 = 10^{-4}$  y  $\varepsilon_2 = 5 \times 10^{-3}$ . Si no se obtiene ninguna de estas convergencias, entonces los estimadores no han convergido a los verdaderos parámetros.

5. Una vez obtenidos los estimadores MVIC, es necesario el cálculo de la matriz de varianzas y covarianzas, la que es igual a la esperanza del

producto exterior de los gradientes evaluados en el estimador MVIC, es decir:  $V(\hat{\theta}) = E[(\nabla \log L)(\nabla \log L)']$ . y en nuestro caso

$$\widehat{V(\hat{\theta})} = \frac{\sum_{i=1}^n [(\nabla \log L_i)(\nabla \log L_i)']}{n} \quad (5.4)$$

donde n es el número de individuos u observaciones y  $\nabla \log L_i$  es el gradiente de la observación i evaluada en  $\hat{\theta}$ .

Los resultados obtenidos se encuentran en detalle en el apéndice A. Las siguientes tablas muestran comparaciones de las estimaciones de LIMDEP 7.0, Stata 7.0, en Ox 3.2 con derivadas numéricas y en Ox 3.2 el procedimiento desarrollado en este trabajo, que utiliza derivadas analíticas.

Ecuación	eq1:era		eq2:eda		eq3:otras	
Paquete	constante	pobre	constante	pobre	constante	pobre
Stata	-1.3367**	0.5325**	-1.0311**	0.4532**	-0.4034**	0.1625**
LIMDEP	-1.3770**	0.5763**	-1.0206**	0.4424**	-0.4020**	0.1603**
Ox (der. num)	-1.3775**	0.5766**	-1.0198**	0.4406**	-0.4009**	0.1589**
Ox (der. ana.)	-1.3764**	0.5770**	-1.0201**	0.4426**	-0.4017**	0.1599**

correlaciones	$\rho_{12}$	$\rho_{13}$	$\rho_{23}$
Stata	0.6455**	0.8735**	0.1166**
LIMDEP	0.6556**	0.1174**	0.1334**
Ox (der. num)	0.6514**	0.1151**	0.1354**
Ox (der. ana.)	0.6550**	0.1142**	0.1327**

\*\* valores significativos al 1%, \* valores significativos al 5%.

Como puede observarse, los resultados son bastante similares entre nuestro nuevo procedimiento y el resto de las estimaciones. El paquete que obtiene resultados más exactos y cercanos a los nuestros es el LIMDEP, ya que existe similitud con LIMDEP hasta el segundo decimal y con Stata hasta el primer decimal. Esto se debe a que Stata utiliza procedimientos más inexactos que LIMDEP (vease Terracol).

El impacto del uso de derivadas analíticas sobre el tiempo es significativo. Nuestra estimación en Ox utilizo 0.11 Horas (aproximadamente 6 minutos),

la estimación con Ox, utilizando derivadas numéricas necesitó 22 minutos, la estimación en LIMDEP 5.85 horas y el procedimiento en Stata 2.4 horas.

Paquete	hrs
Stata	2.3989
LIMDEP	5.8522 <sup>2</sup>
Ox (derivadas numéricas)	0.3650
Ox (derivadas analíticas)	0.1057

Finalmente, los resultados de estas estimaciones reflejan un hecho totalmente intuitivo: El impacto de la pobreza es una causa de enfermedades coyunturales. En lo que concierne al análisis de los efectos marginales, se efectuaron los cálculos de dos tipos básicos.

### Efectos Marginales Incondicionales

Como ya mencionamos anteriormente, los efectos marginales son los gradientes de las probabilidades marginales, dada la estructura de la covariada, éstos deben ser evaluados de manera discreta.

Ecuación	eda	era	otras
$\frac{\Delta F}{\Delta x}$	0.1277**	0.1279**	0.0605**

Utilizando resultados nos permiten concluir que, incondicionalmente, la pobreza afecta de manera más fuerte a los niños menores de 5 años que a otros miembros de la familia.

### Efectos Marginales Conjuntos

En lo que se refiere a los efectos marginales conjuntos, éstos son obtenidos de manera discreta, como el cambio en la probabilidad de pasar de la no pobreza a la pobreza.

$\Pr(eda, era, otras)$	$\frac{\Delta \Pr(\cdot)}{\Delta x}$	$\Pr(eda, era, otras)$	$\frac{\Delta \Pr(\cdot)}{\Delta x}$
$\Pr(eda = 1, era = 1, otras = 1)$	0.0448**	$\Pr(eda = 0, era = 1, otras = 1)$	0.0220**
$\Pr(eda = 1, era = 1, otras = 0)$	0.0432**	$\Pr(eda = 0, era = 1, otras = 0)$	0.0180**
$\Pr(eda = 1, era = 0, otras = 1)$	0.0183**	$\Pr(eda = 0, era = 0, otras = 1)$	-0.0245*
$\Pr(eda = 1, era = 0, otras = 0)$	0.0215**	$\Pr(eda = 0, era = 0, otras = 0)$	-0.1432**

<sup>2</sup>Todos los calculos fueron hechos utilizando LIMDEP versión 7.0 , Stata 7.0 y Ox 3.2. Las estimaciones fueron realizadas con un procesador de 333 Mhz de velocidad.

Los resultados muestran un efecto positivo sobre la probabilidad de adquirir enfermedades de cualquier tipo en el hogar excepto en el caso de la probabilidad  $\Pr(eda = 0, era = 0, otras = 1)$  en la que el impacto de la pobreza no es significativo a un nivel del 1%, y finalmente el impacto de la pobreza reduce la probabilidad de no enfermedades en el hogar.

## 5.2. Un Experimento de Monte Carlo

A fin de obtener un resultado consistente y confirmar el impacto de las derivadas analíticas, propuestas en este trabajo, sobre el tiempo de ajuste, se realizó un experimento de monte carlo utilizando el siguiente Proceso Generador de Datos (PGD).

$$\begin{aligned} y_{i1}^* &= \beta_{11} + \beta_{12}x_{1i} + u_{i1} \\ y_{i2}^* &= \beta_{21} + \beta_{22}x_{2i} + u_{i2} \\ y_{i3}^* &= \beta_{31} + \beta_{32}x_{3i} + u_{i3} \end{aligned} \quad (5.5)$$

donde  $(u_{i1}, u_{i2}, u_{i3})' \sim N_3(0, R)$  y  $R = \begin{bmatrix} 1 & & \\ \rho_{12} & 1 & \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$ ;  $(x_{i1}, x_{i2}, x_{i3}) =$

$N_3\left(\begin{bmatrix} 0,5 \\ -1 \\ 1 \end{bmatrix}, I_3\right)$  y los verdaderos parametros son:  $(\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \beta_{31}, \beta_{32}, \rho_{12}, \rho_{13}, \rho_{23}) = (-0,5, 1, 1, 1, -1, 1, -0,3, 0,2, -0,4)$  y finalmente las variables endógenas observadas son generadas por

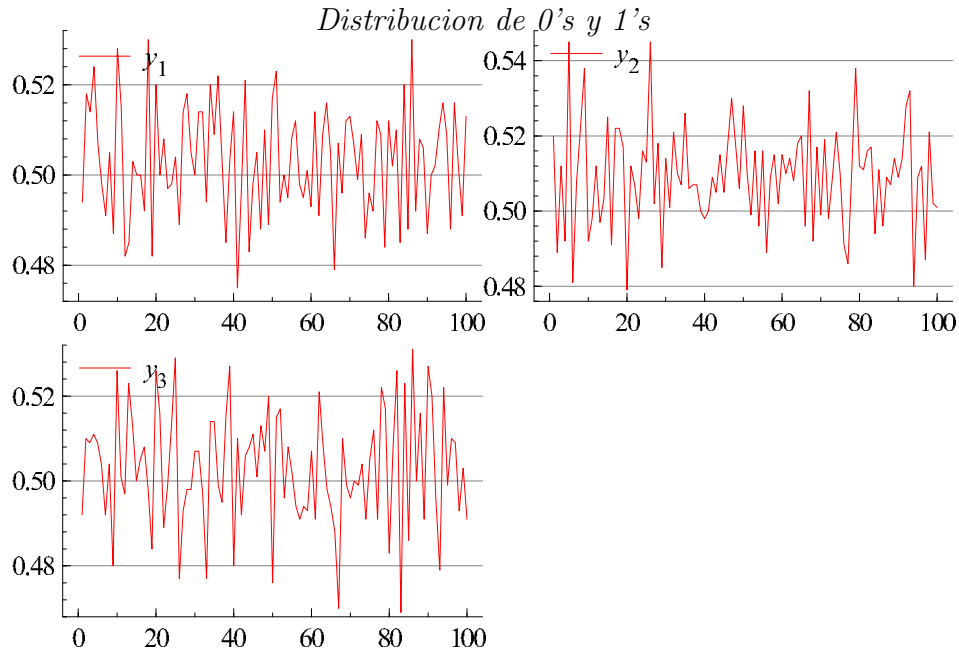
$$y_{ij} = \begin{cases} 1 & \text{if } y_{ij}^* > 0 \\ 0 & \text{if } y_{ij}^* \leq 0 \end{cases}, \quad j = 1, 2, 3; i = 1, \dots, n \quad (5.6)$$

donde  $n$  es el tamaño de la muestra.

El experimento se realizó utilizando un tamaño muestral de 1000 observaciones y fue replicado 100 veces utilizando la técnica de reducción de varianza, denominada "variables antitéticas" (se reutilizan los números aleatorios generados, con el signo opuesto, de forma que se elimina la variabilidad de las estimaciones, vease Davidson y Mackinon pag. 744-747). En cada paso, los estimadores MVIC fueron calculados en Ox con derivadas analíticas y

numéricas; la estimación utilizó los pasos 1 a 5 descritos en la sección anterior. Se aplicó el método GHK con 100 iteraciones para el cálculo de cada probabilidad.

En los gráficos siguientes se ilustran la proporción de 1's en cada muestra.



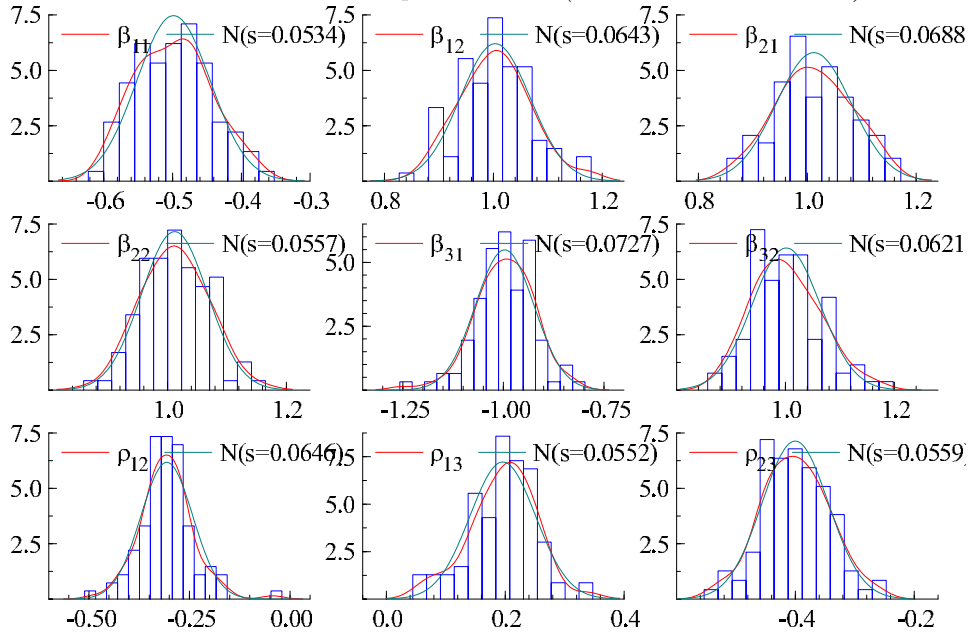
Como se puede observar, la composición de 0's y 1's es proporcional en la mayor parte de las muestras, solo  $y_2$  posee un nivel una proporción mayor de 1's.

En lo referente a los tiempos de estimación (en segundos), cada muestra fue estimada con derivadas numéricas y luego con derivadas analíticas. Los resultados muestran el uso de derivadas analíticas permite obtener las estimaciones tres veces más rápido, en promedio, que utilizando los procedimientos actuales.

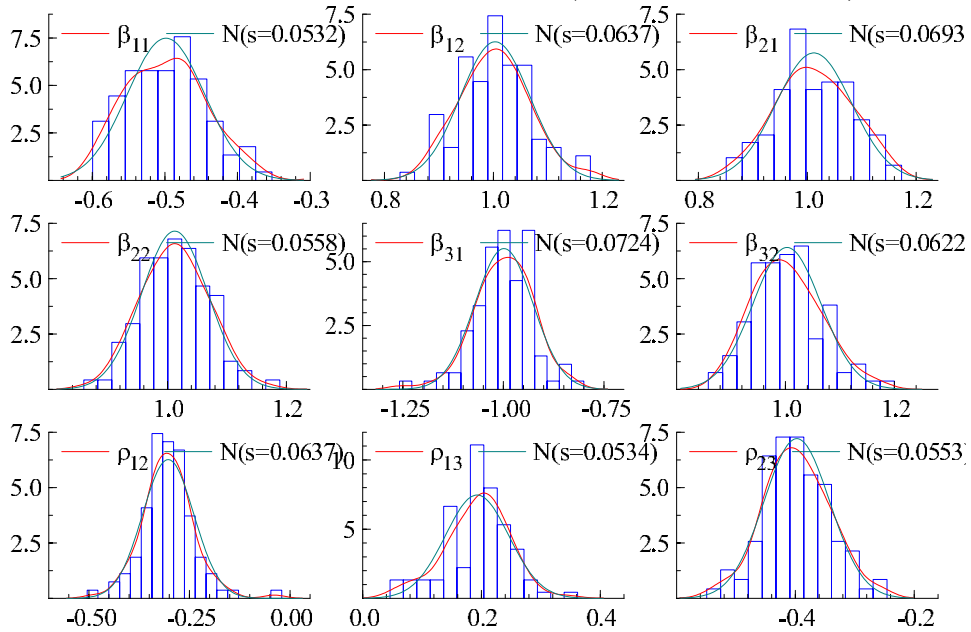
	Derivadas Numéricas	Derivadas Analíticas
tiempo promedio	211.2442 seg	69.2553 seg

En cuanto a los estimadores, las distribuciones empíricas de ambos procedimientos de estimación son bastante cercanas entre sí, lo que indica similitud en las distribuciones.

*Distribuciones de los parámetros (derivadas numéricas)*



*Distribuciones de los parámetros (derivadas analíticas)*



## Resultados del Experimento de Monte Carlo

	$\beta_{11}$	$\beta_{12}$	$\beta_{21}$	$\beta_{22}$	$\beta_{31}$	$\beta_{32}$	$\rho_{12}$	$\rho_{13}$	$\rho_{23}$
VP	-0.5000	1.0000	1.0000	1.0000	-1.0000	1.0000	-0.3000	0.2000	-0.4000
P (num)	-0.4985	1.0033	1.0121	1.0122	-0.9986	1.0028	-0.3028	0.1926	-0.3976
P (ana)	-0.4985	1.0034	1.0119	1.0119	-0.9969	1.0013	-0.3063	0.1959	-0.4004
DE (num)	0.0532	0.0637	0.0693	0.0558	0.0724	0.0622	0.0637	0.0534	0.0553
DE (ana)	0.0534	0.0643	0.0688	0.0557	0.0727	0.0621	0.0646	0.0552	0.0559
Q(0.005) (num)	-0.5965	0.8692	0.8583	0.8829	-1.2231	0.8799	-0.4586	0.0544	-0.5308
Q(0.005) (ana)	-0.5994	0.8680	0.8591	0.8840	-1.2211	0.8767	-0.4616	0.0557	-0.5306
M (num)	-0.4949	1.0008	1.0077	1.0117	-0.9929	0.9981	-0.3018	0.1961	-0.3984
M (ana)	-0.4961	1.0013	1.0080	1.0118	-0.9928	0.9963	-0.3062	0.2005	-0.3990
Q(0.995) (num)	-0.3771	1.1756	1.1590	1.1502	-0.8356	1.1671	-0.0935	0.3278	-0.2580
Q(0.995) (ana)	-0.3781	1.1766	1.1563	1.1472	-0.8338	1.1648	-0.1015	0.3386	-0.2652

VP: Valor Poblacional; P: Valor Promedio, DE: Desviación estándar, Q(0.005): Quantil 0.5 centésimo, Q(0.995) Quantil 99.5 centésimo, M : Mediana: (num): Resultado obtenido utilizando derivadas numéricas; (ana): Resultado obtenido utilizando derivadas analíticas.

Los resultados evidencian los siguientes hechos:

1. El uso de derivadas analíticas reduce el tiempo de estimación en aproximadamente tres veces.
2. El uso de las derivadas analíticas es consistente dado que convergen, en distribución, a los resultados obtenidos con los procedimientos usuales (derivadas numéricas).



# Capítulo 6

## Conclusiones y Recomendaciones

Varias interesantes conclusiones emergen de este trabajo:

Los modelos PMV poseen estructuras complicadas en su diseño y precisan de mayor análisis estadístico, sobre todo en las condiciones de identificación. La función de verosimilitud, asociada a cada modelo, es una expresión compleja que contiene integrales de grado superior que no poseen una expresión analítica y para las estimaciones, las condiciones de primer orden se calculan de forma numérica debido a la ausencia de formas analíticas.

Este trabajo se constituye en una contribución a la estimación de estos modelos, ya que presenta las expresiones analíticas de las derivadas de primer grado de la función de verosimilitud, respecto a los parámetros y correlaciones. Estas expresiones permiten, en primer lugar, obtener resultados más precisos ya que son formas exactas y, en segundo lugar, mejoran el tiempo de estimación ya que reducen el número de cálculos necesarios para obtener los estimadores MVIC.

Los diferentes efectos marginales emergentes de las distintas probabilidades del modelo PMV, que permiten completar el análisis de los problemas económicos modelizados, han podido ser estimados gracias a los resultados obtenidos en este trabajo. En efecto, anteriormente, estos efectos marginales requerían de programas altamente sofisticados, complejos y de convergencia lenta para ser obtenidos y, a excepción del LIMDEP que tiene la opción para obtener los efectos marginales condicionales, ningún paquete informático especializado en econometría ofrece estos estadísticos.

Finalmente, se propone continuar la investigación de este tema, en dos

áreas importantes: El problema de la identificación de los modelos PMV y el estudio de nuevos y mejores métodos de estimación de las probabilidades de los modelos PMV con grado superior a dos.

# Bibliografía

- [1] Amemiya, T. (1974), "*Multivariate Regression and Simultaneous Equation Models when the Dependent Variables are Truncated Normal*", *Econometrica*. 42; 999 - 1012.
- [2] Amemiya, T. (1977), "*The estimation of a simultaneous equations generalized probit model*", *Econometrica*, 46; 1193 - 1205.
- [3] Broyden, C. G. (1970), "*The convergence of a class of double rank minimization algorithms, partes I y II*", *Journal of the Institute of Mathematics and its Applications* 6, 76-90.
- [4] Chib, Siddharta y Edward Greenberg (1998), "*Analysis of multivariate probit models*", *Biometrika* 85-2; 347-361
- [5] Dagenais, Marcel G. (1997), "*A Simultaneous Probit Model*", Documento no Publicado.
- [6] Dagenais, Marcel G. (1999), "*Inconsistency of a proposed nonlinear instrumental variables estimator for probit and logit models with endogenous regressors.*", *Economic Letters*.
- [7] Davidson, James (2000), "*Econometric Theory*", Blackwell Publishers, United Kingdom..
- [8] Davidson, Russell , y James G. Mackinnon (1993), "*Estimation and Inference in Econometrics*", Oxford University Press, New York.
- [9] Deaton, Angus (1998), "*The Analysis of Household surveys: a microeconomic approach to development policy*", The John Hopkins University Press.
- [10] Doornik, Jurgen A.(2002). "*Ox Appendices*", versión no publicada.

- [11] Doornik, Jurgen A., y David F. Hendry (2001), *“GiveWin: An Interface to Empirical Modelling”*, Timberlake Consultants Press., United Kingdom.
- [12] Doornik, Jurgen A., David F. Hendry, y Neil Shephard (2001), *“Computationally-intensive Econometrics using a Distributed Matrix-programming Language”*, Artículo no publicado. Introduction to Ox
- [13] Doornik, Jurgen A. y Marius Ooms (2001), *“Introduction to Ox”*, Timberlake Consultants Press., United Kingdom.
- [14] Greene, William H. (2000), *“Econometric Analysis”*, Prentice Hall ( 4ta. Edición)
- [15] Greene, William H. (1998), *“Gender Economics in Liberal arts College: Further Results”*, Journal of Economic Education, 29(4)
- [16] Greene, William H. (1996), *“Marginal Effects in the Bivariate probit Model”*, Working Paper, EC-96-11, Department of Economics, Stern School of Business, New York University.
- [17] Hajivassiliou, V. A. (1993), *“Simulation Estimation Methods for Limited Dependent Variable Models”*, en *“Handbook of Statistics”*, Vol 11, editado por G. S. Maddala , C. R. Rao y Vinod, H. D., Elsevier Science Publishers B. V.
- [18] Hajivassiliou, V. A. y P. A. Ruud (1994), *“Classical estimation methods for LDV models using simulation”*, en *“Handbook of Econometrics”*, Vol. II., Editado por R. F. Engle y D. L. McFadden, Amsterdam: North-Holland.
- [19] Iwarere, Jide L., y Williams, John E.(1991), *“A Micro-Market Analysis of Tenure Choices Using The Logit Model ”*, The Journal of Real State 6(1), Fannie Mae Foundation.
- [20] Lee, Lung-Fei (1990), *“Simultaneous Equations with Discrete and Censored Variables”*, en *“Structural Analysis of Discrete Data with Econometric Applications”*, Editado por Manski, Charles F. y McFadden, Daniel L., MIT Press.

- [21] Long, Scott J. (1997), *“Regression Models for Categorical and Limited Dependent Variables”*, Sage Publications.
- [22] Maddala, G. S. (1983), *“Limited-Dependent and Qualitative Variables in Econometrics”*, Cambridge University Press.
- [23] McFadden, Daniel L. (1984)., *“Econometric Analysis of Qualitative Response Models”*, en *“Handbook of Econometrics”*, Vol. II., Editado por Z. Griliches y M. D. Intriligator, Amsterdam: North-Holland.
- [24] Pohlmeier, W. (1992), *“On the Simultaneity of Innovations and Market Structure”*, *Empirical Economics*, 17; 253-72.
- [25] Poirier, Dale J. (1995), *“Intermediate Statistics and Econometrics”*, MIT Press.
- [26] Quandt, Richard E. (1983), *“Computational Problems and Methods”*, en *“Handbook of Econometrics”*, Vol. I., Editado por Z. Griliches y M. D. Intriligator, Amsterdam: North-Holland.
- [27] Smith, Peter (1981), *“Constraints on the parameters in simultaneous tobit and probit models ”*, en *“Structural Analysis of Discrete Data with Econometric Applications”*, Editado por: Manski, C. F. y MacFadden, D., MIT Press.
- [28] Stata (2001), *“Stata Programing Manual Release 7”*, College Station, TX: Stata Press.
- [29] Stata (2001), Manuals 1- 4, *“Stata Reference Manuals Release 7”*. College Station, TX: Stata Press.
- [30] Terracol, Antoine (2001), *“Triprobit and the GHK simulator: a short note”*, documento no publicado.
- [31] Wilde, Joachim (2000), *“Identification of multiple equation probit models with endogenous dummy regressors”*, *Economic Letters* 69; 309-312.
- [32] Zellner, A. Y T. Lee (1965), *“Joint estimation on relationships involving discrete random variables”*, *Econometrica* 33; 383-94

# Apéndice A

## Resultados de la estimación

Estimación en STATA (triprobit por Antoine Terracol)

Log likelihood = -8228.953

Number of obs = 5333  
Wald chi2(3) = 211.31  
Prob > chi2 = 0.0000

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----						
enf1						
p0	.5325486	.0427478	12.46	0.000	.4487646	.6163327
_cons	-1.336726	.0351448	-38.03	0.000	-1.405608	-1.267843
-----						
enf2						
p0	.4531767	.0401441	11.29	0.000	.3744958	.5318576
_cons	-1.031078	.032517	-31.71	0.000	-1.09481	-.9673462
-----						
enf3						
p0	.1625134	.0357058	4.55	0.000	.0925314	.2324954
_cons	-.4034004	.0277758	-14.52	0.000	-.4578401	-.3489608
-----						
athrho12						
_cons	.7675853	.030955	24.80	0.000	.7069147	.8282559
-----						
athrho13						
_cons	.0875787	.0234436	3.74	0.000	.0416301	.1335272
-----						
athrho23						
_cons	.1171502	.0228946	5.12	0.000	.0722776	.1620227
-----						

rho12= .64552315 Std. Err.= .01805603 z= 35.751116 Pr>|z|= 0  
rho13= .08735543 Std. Err.= .02326468 z= 3.7548524 Pr>|z|= .00017344  
rho23= .11661718 Std. Err.= .02258322 z= 5.1638872 Pr>|z|= 2.419e-07

LR test of rho12=rho13=rho23=0: chi2(3) = 761.83323 Prob > chi2 = 8.193e-165

-----

Time elapsed 2.398888 hrs.

Estimación en LIMDEP (Mprobit por Willian Greene)

```

+-----+
| Multivariate Probit Model: 3 equations. |
| Maximum Likelihood Estimates           |
| Dependent variable                     MVProbit |
| Weighting variable                     ONE       |
| Number of observations                  5333     |
| Iterations completed                   13       |
| Log likelihood function                -8226.203 |
| Replications for simulated probs. = 100 |
+-----+

```

```

+-----+-----+-----+-----+-----+
|Variable | Coefficient | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+-----+-----+-----+-----+-----+
      Index function for ENF1
Constant -1.377040328      .38568056E-01  -35.704  .0000
PO       .5762884992      .45823163E-01   12.576  .0000  .59703731
      Index function for ENF2
Constant -1.020638110      .32666839E-01  -31.244  .0000
PO       .4423940776      .40256970E-01   10.989  .0000  .59703731
      Index function for ENF3
Constant -.4020130729      .27853991E-01  -14.433  .0000
PO       .1603553394      .35771937E-01   4.483   .0000  .59703731
      Correlation coefficients
R(01,02) .6555887482      .18258250E-01   35.906  .0000
R(01,03) .1174061569      .26084454E-01   4.501   .0000
R(02,03) .1334310637      .23925161E-01   5.577   .0000

```

Elapsed time: 5 hours, 51 minutes, 7.77 seconds.



Estimación Ox usando Derivadas Numéricas

Strong convergence using BFGS method with first analytical derivates

Covariance matrix estimation : BHHH

Log-likelihood : -8227.82, BIC :-29233.2

no. of observations : 5333, no. of parameters : 9

---

eq1 : enf1

	coef.	std.	z-sta	p-val
Constant	-1.3775	0.0385	-35.8200	0.0000
p0	0.5766	0.0457	12.6198	0.0000

eq2 : enf2

	coef.	std.	z-sta	p-val
Constant	-1.0198	0.0327	-31.2009	0.0000
p0	0.4406	0.0402	10.9500	0.0000

eq3 : enf3

	coef.	std.	z-sta	p-val
Constant	-0.4009	0.0278	-14.3987	0.0000
p0	0.1589	0.0357	4.4471	0.0000

Residual correlations

	coef.	std.	z-sta	p-val
rho12	0.6514	0.0184	35.3259	0.0000
rho13	0.1151	0.0262	4.4012	0.0000
rho23	0.1354	0.0239	5.6548	0.0000

---

-----Time Elapsed = 21:54.10-----

Estimación Ox usando Derivadas Analíticas

Strong convergence using BFGS method with first analytical derivatives  
 Covariance matrix estimation : BHHH  
 Log-likelihood : -8227.85, BIC :-29233.1  
 no. of observations : 5333, no. of parameters : 9

---

eq1 : enf1

	coef.	std.	z-sta	p-val
Constant	-1.3764	0.0384	-35.8060	0.0000
p0	0.5770	0.0457	12.6360	0.0000

eq2 : enf2

	coef.	std.	z-sta	p-val
Constant	-1.0201	0.0327	-31.2212	0.0000
p0	0.4426	0.0402	11.0020	0.0000

eq3 : enf3

	coef.	std.	z-sta	p-val
Constant	-0.4017	0.0278	-14.4273	0.0000
p0	0.1599	0.0357	4.4750	0.0000

Residual correlations

	coef.	std.	z-sta	p-val
rho12	0.6550	0.0183	35.7728	0.0000
rho13	0.1142	0.0261	4.3680	0.0000
rho23	0.1327	0.0240	5.5389	0.0000

---

-----Time Elapsed = 6:40.84-----

Resultados adicionales Ox Derivadas Analíticas

Marginal Effects of Inconditional distributions: at Median dF/dX : 0 -> 1

Constant	p0			
1.0000	1.0000			
eq1 : enf1				
	dF/dx.	std.	z-sta	p-val
p0	0.1277	0.0093	13.7001	0.0000
eq2 : enf2				
	dF/dx.	std.	z-sta	p-val
p0	0.1279	0.0111	11.5422	0.0000
eq3 : enf3				
	dF/dx.	std.	z-sta	p-val
p0	0.0605	0.0134	4.5070	0.0000

Marginal Effects of Joint Distributions: at Median dF/dX : 0 -> 1

Marginal Effects at :

Constant	p0			
1.0000	1.0000			
Prob(enf1 = 1, enf2 = 1, enf3 = 1)				
	dF/dx.	std.	z-sta	p-val
p0	0.0448	0.0033	13.4956	0.0000
Prob(enf1 = 1, enf2 = 1, enf3 = 0)				
	dF/dx.	std.	z-sta	p-val
p0	0.0432	0.0036	11.8541	0.0000
Prob(enf1 = 1, enf2 = 1, enf3 = 1)				
	dF/dx.	std.	z-sta	p-val
p0	0.0183	0.0022	8.1516	0.0000
Prob(enf1 = 1, enf2 = 0, enf3 = 0)				
	dF/dx.	std.	z-sta	p-val
p0	0.0215	0.0031	7.0091	0.0000
Prob(enf1 = 0, enf2 = 1, enf3 = 1)				
	dF/dx.	std.	z-sta	p-val
p0	0.0220	0.0038	5.7092	0.0000
Prob(enf1 = 0, enf2 = 1, enf3 = 0)				
	dF/dx.	std.	z-sta	p-val
p0	0.0180	0.0050	3.5730	0.0004
Prob(enf1 = 0, enf2 = 0, enf3 = 1)				
	dF/dx.	std.	z-sta	p-val

p0	-0.0245	0.0106	-2.3039	0.0212
Prob(enf1 = 0,enf2 = 0,enf3 = 0)				
	dF/dx.	std.	z-sta	p-val
p0	-0.1432	0.0125	-11.4538	0.0000